DOCUMENT RESUME

ED 477 296 SP 040 879

AUTHOR	Darling-Hammond, Linda
TITLE	The Research and Rhetoric on Teacher Certification: A Response to "Teacher Certification Reconsidered."
PUB DATE	2001-10-00
NOTE	72p.
AVAILABLE FROM	For full text: http://www.nctaf.org/publications/abell_response.pdf.
PUB TYPE	Opinion Papers (120)
EDRS PRICE	EDRS Price MF01/PC03 Plus Postage.
DESCRIPTORS	Academic Achievement; Educational Research; Elementary Secondary Education; Higher Education; *Inner City; Minority Group Children ; Poverty; *Preservice Teacher Education; Private Schools; Research Methodology; *Teacher Certification; Teacher Competencies; Teacher Effectiveness; Teacher Influence; Teacher Persistence; *Teacher Qualifications; Teacher Recruitment; Urban Schools
IDENTIFIERS	Baltimore City Public Schools MD

ABSTRACT

This paper examines Kate Walsh's Abell Foundation report, which purports to prove there is no credible research supporting the use of teacher certification as a regulatory barrier to teaching and argues against reforms that would strengthen incentives to bring qualified teaches to inner city schools. This paper discusses inaccuracies in Walsh's account, actual findings of many studies it purports to review, and findings from other studies Walsh's report ignores. Five major issues this paper addresses are: evidence about student learning in reading and other areas that is ignored; unfounded claims; misrepresentations of research; methodological issues and double standards in using research; and illogical policy conclusions. This paper asserts that Walsh has dismissed or misreported much of the existing evidence base in order to argue that teacher education makes no difference to teacher performance or student learning and that students would be better off without state efforts to regulate entry into teaching or to provide support for teachers' learning. While Walsh's proposal is couched as the elimination of barriers to teaching, evidence suggests that lack of preparation actually contributes to high attrition rates and thereby becomes a disincentive to long-term teaching commitments and the creation of a stable, high ability teaching force (which leads to lower learning levels). (Contains 89 references.) (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

SP040879
6/4/02.

ED 477 296

The Research and Rhetoric on Teacher Certification: A Response to "Teacher Certification Reconsidered"

Linda Darling-Hammond

Charles E. Ducommun Professor of Education, Stanford University
Executive Director, National Commission on Teaching and America's Future

October 15, 2001

SP040879

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

C.H. Simon

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

BEST COPY AVAILABLE

Table of Contents

What are the Arguments?	4
Evidence Ignored	6
Unfounded Claims	7
The Effectiveness of Certified and Uncertified Teachers	8
Evidence about Preservice Teacher Education	20
The Influence of Verbal Ability on Teacher Effectiveness	25
The Academic Ability of Teachers who Lack Certification	36
The Private School Argument	39
Other Misrepresentations of Research Findings	43
Methodological Issues	51
Study Size and Design	52
Level of Aggregation	53
Measures for Assessing Teaching Performance	55
Age of Studies	55
Publication Venue	56
Who is Affected by this Debate?	57
Illogical Policy Conclusions	58
Conclusion	60
References	63

The Research and Rhetoric on Teacher Certification:
A Response to "Teacher Certification Reconsidered"

L. Darling-Hammond

In a stunning exercise in misrepresentation, Kate Walsh has written a paper for the Abell Foundation[1] that purports to prove that there is "no credible research that supports the use of teacher certification as a regulatory barrier to teaching" (p. 5). She contends that "educators, policymakers, the media, and the public mistakenly equate teacher quality with teacher certification" (p. 1). Her agenda is to argue against reforms that would strengthen incentives to bring qualified teachers to inner city schools in cities like Baltimore, Maryland, a prime target of the paper. A related agenda is to argue against Maryland's efforts to strengthen teacher preparation and certification and to expand incentives for improving the supply of certified teachers in schools across the state. A final agenda is to rekindle support for the Resident Teacher Program in Baltimore, a program that has been a revolving door of under-prepared teachers into and out of Baltimore Public Schools, recently targeted for discontinuation by the new superintendent of schools because of its high attrition rates and poor outcomes for children.

Walsh complains that efforts to improve education for poor and minority children in Baltimore by the state and local superintendents of schools and by local advocacy organizations wrongly seek to secure more fully certified teachers for their schools. She cites as wrong-headed newspaper articles raising concerns, for example, that: "Least prepared teachers are at worst city schools: One-third lack basic credentials for certification," (p. 1). She ridicules the efforts of a Baltimore community group that released a study which "bemoaned the fact that more uncertified teachers were teaching in the city's high-poverty, predominantly African-American

[1] "Teacher Certification Reconsidered: Stumbling for Quality," sponsored by the Abell Foundation, is published through the Abell Foundation website.

schools than the city's whiter, more affluent schools," (p. 2) suggesting that their efforts are misguided. Walsh also strongly criticizes recent efforts to strengthen Maryland's certification requirements by requiring courses in the teaching of reading for all teachers, characterizing them as additional "barriers" to the ability to teach.

Walsh proposes that Maryland should 1) "eliminate the coursework requirements for teacher certification" and require only a bachelor's degree and a passing score on an appropriate teacher's exam; 2) "report the average verbal ability score of teachers in each school district and of teacher candidates graduating from the State's schools of education" (but presumably not those who come in without training for teaching); and 3) "devolve its responsibility for teacher qualification and selection to its 24 public school districts," delegating all hiring authority to individual school principals (pp. vii-viii). In suggesting that this is the answer to the problem of recruitment for the schools serving minority and poor children, Walsh ignores the fact that, even if all principals had infinite information at their disposal about the likely effectiveness of teachers and made wise, fully informed choices (two assumptions that have been challenged by some research on teacher selection practices), principals do not control the major levers for addressing the problems of unequal supply: unequal district revenues, noncompetitive teacher salary levels, and the policies that govern recruitment and preparation that would allow them to seek out and hire the individuals they might most want to recruit.

Eliminating certification requirements would eliminate pressures for competitive wages or recruitment incentives for teachers, since an open marketplace in a resource-constrained public sector could resolve shortages by lowering standards. In addition, eliminating certification requirements would eliminate evidence about disparities in students' opportunities to learn, for if there are no minimum standards, there will be no evidence of differences in the

extent to which they have been achieved by teachers working with different groups of students. This would in turn reduce pressures for the creation of policies to rectify these inequities. Finally, eliminating such standards would remove the mechanisms states have been developing and improving to be sure that teachers know their content well, know how to teach the content to students, know how to teach fundamental skills like reading, and have the ability to meet the special needs of learners who may have learning disabilities that require distinct teaching strategies, whose first language is not English, or who simply struggle with certain kinds of academic tasks and need diagnostic assistance.

In order to make a case for this agenda, Walsh attacks all research that has found relationships between teachers' preparation and their measured effectiveness, including students' achievement. She goes on to characterize much of the education research as "flawed, sloppy, aged and sometimes academically dishonest" (p. 13), a characterization that more aptly describes her paper, which consistently misrepresents the statements of researchers, the findings of studies, and the evidence base for her claims.

All studies have limitations, and some are too problematic to be relied upon, including a number that Walsh relies upon for her own assertions. However, Walsh's paper, which is littered with dozens of inaccuracies, misstatements, and misrepresentations, sheds little light on them or their implications for the research base on teacher education and certification. In what follows I discuss the inaccuracies in Walsh's account, the actual findings of many of the studies she purports to review, and the findings of other studies she chooses to ignore, as well as the implications of her proposals for teachers, their knowledge, and the students they teach.

What are the Arguments?

The Abell Foundation report admits that teacher qualifications make a difference but it also tries to make a case that "the backgrounds and attributes characterizing effective teachers are more likely to be found outside the domain of schools of education. The teacher attribute found consistently to be most related to raising student achievement is verbal ability.... usually measured by short vocabulary tests..." (p. v). Later in the report, Walsh suggests that subject matter knowledge may be an additional criterion for hiring secondary teachers, but not for elementary teachers. (Walsh objects to the state requirements regarding content coursework for elementary teachers, since many who want to enter through the alternative Resident Teacher program have had trouble meeting these requirements.) Walsh then tries to dismiss all studies that find evidence that knowledge about teaching also makes a different for teacher performance, or to claim that studies finding positive effects of teacher education or certification are either too old, too small, too highly aggregated, or dependent on evidence about teacher performance other than student achievement or are not really about certification after all, even if their authors say they are. She often does this by misrepresenting the studies' actual methods and findings, as I detail below.

While there are legitimate concerns to be raised about various studies in the literature – on all sides of the question – this paper does not shed much light on them. A thorough review of the quality and accurately portrayed findings of the several bodies of research that bear on this question would be a real service to this field. Unfortunately, this document's extensive inaccuracies and misunderstanding of the research it cites make it of little use in this regard.

In what follows, I address five major issues regarding the Abell report and the research base on teaching and teacher education:

1) **Evidence Ignored.** Evidence about student learning in reading and other areas documents the need for teachers to have professional knowledge that includes and extends beyond subject matter knowledge. The Abell Foundation report does not consider this evidence or answer the question of how teachers are to acquire this knowledge if they are not professionally prepared.

2) **Unfounded Claims.** No evidence supports Walsh's claim that either verbal ability or subject matter knowledge alone makes teachers effective. She lacks supporting evidence – and fails to consider contradictory evidence – for her claims about the relative effectiveness of certified and uncertified teachers, the outcomes of teacher education, the primacy of verbal ability as the most important measure of teaching, the effectiveness of private and public schools and the preparation of their teachers, and the attributes of individuals who enter teaching without certification.

3) **Misrepresentations of Research.** Walsh's claim that she has reviewed 100 to 200 studies cited in support of teacher education and found that "none of them holds up to scrutiny" is not true. A large number of the studies cited as relevant to the question of teacher education effects are not reviewed at all in Walsh's paper. Many of those reviewed are badly misrepresented, including inaccurate statements about their methods and findings, false claims about their authors' views, and distortions of their data and conclusions. Many are not reviewed for their methods and findings, but are dismissed because of their sample size, age, dependent variable, or publication venue – unless Walsh likes one of the findings, in which case she uses the study, sometimes after already having dismissed it. Even the studies that Walsh says she reviewed are missing from the appendix of the report, where she refers readers for evidence.[2]

4) **Methodological Issues and Double Standards in Using Research.** Walsh misunderstands some fundamental research design issues, including the difference between experimental and correlational studies and the interpretation of research conducted at different levels of aggregation. In her effort to make the evidence base about teacher education disappear, Walsh eliminates from consideration studies that have been cited regarding the contributions of various measures of teacher qualifications to teacher effectiveness if they have small sample sizes, if they were published more than 20 years ago, or if they were published as dissertations, technical reports, or conference papers rather than in peer-reviewed journals. She also eliminates all studies that use measures of teacher effectiveness other than student

[2] See, for example, footnote 18 on p. 13 where Walsh refers readers to Appendix B for analysis of six studies, only two of which (Guyton & Farokhi, 1987; Monk, 1994) are actually included there. Appendix B of the published version of Walsh's report includes only 14 of 192 studies originally included in her draft of July 23, 2001 and does not include most of the key studies on the topic. A more complete appendix was later added to the Abell Foundation website. Readers must consult with that document if they want to see Walsh's comments on studies. The comments are often random observations about a study and are rarely complete enough to constitute a review.

BEST COPY AVAILABLE

achievement (e.g. supervisors' ratings of performance, researchers' observation-based measures of teacher practice). There are legitimate issues associated with the sample size, age, quality assurance, and measurement that warrant discussion (see below). However, as a blanket means of eliminating evidence from consideration, this strategy is problematic, as Walsh's frequent citations of studies that fail to meet her own criteria suggest.

5) **Illogical Policy Conclusions.** While it is clear that teacher certification systems are not perfect and there are many weak teacher education programs, points that I have frequently made in my own research, it does not follow that the response to these problems should be to eliminate expectations for teachers to acquire the knowledge they need to teach students effectively. The more appropriate policy response is to improve the quality of teacher education – a process that has been underway with important results in a number of states including Maryland, and one that rests on the processes of accreditation and certification that provide policymakers with levers for change and improvement.

Evidence Ignored

While the Abell Foundation report claims that teachers do not need professional knowledge in order to teach, the field has been moving rapidly to codify the ways in which teaching knowledge makes a difference in student learning. For example, the National Reading Panel of the National Institute of Child Health and Human Development last year published a major review of carefully controlled research which found that children's reading achievement is improved by systematic teaching of phonemic awareness, guided repeated oral reading, direct and indirect vocabulary instruction with careful attention to readers' needs, and a combination of reading comprehension techniques that include metacognitive strategies.

The report notes that teacher education is critical to the success of reading instruction with respect to both instruction in phonemic awareness and more complex comprehension skills:

> Knowing that all phonics programs are not the same brings with it the implication that teachers must themselves be educated about how to evaluate different programs to determine which ones are based on strong evidence and how they can most effectively use these programs in their own classrooms. It is therefore important that teachers be

> provided with evidence-based preservice training and ongoing inservice training to select (or develop) and implement the most appropriate phonics instruction effectively. (p. 11)
>
> Teaching reading comprehension strategies to students at all grade levels is complex. Teachers not only must have a firm grasp of the content presented in the text, but also must have substantial knowledge of the strategies themselves, of which strategies are most effective for different students and types of content and of how best to teach and model strategy use.... (Data from the studies reviewed on teacher training) indicated clearly that in order for teachers to use strategies effectively, extensive formal instruction in reading comprehension is necessary, preferably beginning as early as pre-service (National Reading Panel, 2000, pp. 15-16).

Similar insights in our understanding of how to develop student proficiency in mathematics and science, and how to develop teachers' skills for doing so, have recently emerged. For example, recent analyses of the National Assessment of Educational Progress (NAEP) which control for a number of measures of school inputs and student characteristics have found that students whose teachers have majored in mathematics or mathematics education, who have had more pre- or in-service training in how to work with diverse student populations and more training in how to develop higher-order thinking skills, and who engage in more hands-on learning do better on the NAEP mathematics assessments. Similarly, students whose teachers have majored in science or science education and who have had more pre- or in-service training in how to develop laboratory skills and who engage in more hands-on learning do better on the NAEP science assessments (Wenglinsky, 2000).

It is ironic that just as the field is learning more about how to prepare teachers to teach children effectively, the Abell Foundation suggests that we truncate teacher education and end the certification policies that would encourage and enable teachers to acquire this knowledge – or at least that we do so for the children of the poor, who also attend school in districts with minimal resources for professional development. The unanswered question is, How are teachers

to learn what is known about how to teach well if there are no expectations, incentives, or supports for them to do so?

Unfounded Claims

While ignoring these serious questions, Walsh makes a number of claims that are not supported either by the research she presents or by other evidence in the field. These include at least the following:

- New teachers who are certified do not produce greater student gains than new teachers who are not certified.
- There is little evidence that the content and skills taught in preservice education coursework is (*sic*) either retained or effective.
- Verbal ability and subject matter alone are sufficient to produce effective teachers.
- Private schools do not hire certified teachers and they are more effective than public schools.
- Individuals with higher academic ability will be recruited to teaching if certification standards eliminated.

The Effectiveness of Certified and Uncertified Teachers

For her proposition that "new teachers who are certified do not produce greater student gains than new teachers who are not certified," Walsh cites seven studies, at least six of which do not provide support for this proposition, and five of which actually provide contradictory evidence to her claim.[3] The seventh, an unpublished study, could not be retrieved in the brief period of time available for this reply. Three of the studies (Bliss, 1992; Stoddart, 1992; Lutz & Hutton, 1989) include no data on student achievement at all, although Walsh elsewhere dismisses all other studies that do not use student achievement data as the dependent variable. Five of the studies actually deal with alternatively certified rather than uncertified teachers – that

is, teachers who had undertaken teacher education at the post-baccalaureate level in university- or school district-based programs that rearrange the way teacher education is delivered. The findings across the studies are mixed, but none of them shows that uncertified teachers do as well as certified teachers, and one of them shows that this is clearly not true. Several of the studies point to the value of teacher education: The more positive findings are found for the alternatives that provide more complete preparation.

1. **Bliss (1992)** wrote about the Connecticut alternative certification program, a two-year training model which the author notes features "a significantly longer period of training than in any other alternate route program" in existence at that time (p. 52). This report does not examine uncertified teachers, nor does it meet Walsh's criteria for inclusion in a review of literature, because it includes no data about teacher effectiveness as gauged by student achievement measures. Bliss notes that most recruits reported their initial training to be helpful, and she briefly mentions results from another researcher's survey of recruits' supervisors which suggested mixed reviews of their performance: 33 percent of supervisors said that the alternate route teachers were weaker than others in classroom management (presumably, then, 67 percent said they were not weaker than others in this area), while 38 percent said they were stronger than others in teaching skills (and 62 percent presumably said they were not stronger than others in this area).

2. **Stoddart (1992)** reports on the subject matter qualifications and attrition rates of recruits to the Los Angeles Teacher Trainee Program, also a two-year training model. She found that content qualifications were comparable to those of traditionally trained recruits, except for math recruits, who had lower GPAs than traditionally trained teachers, and that attrition rates for those who entered were relatively low in the first two years but higher than national rates after 5

---

[3] Three of the seven studies cited by Walsh (Bradshaw & Hawk, 1996; Stoddart, 1992; and Bliss, 1992) are not listed in her "References" section; one of them (Stoddart, 1992) also does not appear in the separately-published

years.[4] Results cited by Stoddart from other studies about the observed practices of these teachers in comparison with university-trained teachers produced mixed results: university-trained English teachers appeared more skillful than alternate route teachers, but the levels of skill appeared lower for mathematics teachers from both groups.

3. **Lutz and Hutton (1989)** compared the demographic characteristics, attitudes, certification test scores, and opinions of Dallas Public Schools' alternative certification (AC) recruits with other first year teachers in the district. Like the other studies noted above, this study did not examine student achievement gains of the recruits' students. The program provides summer training to recruits and then places them in mentored internships during the school year while they are completing other coursework. The study found many similarities but some differences between AC recruits and other first year teachers, including significantly lower rates of expected long-term continuation in teaching for the AC recruits (40% vs. 72% for other first year teachers). They also examined supervisors' perceptions of recruits – a measure that Walsh argues should eliminate other studies from consideration. These were positive for the 54% of the pool (59 out of 110) defined as "successful" interns in the study – those who completed the intern year without dropping out (10%) or being held back for another year or more due to 'deficiencies' in coursework or other areas of performance (36%). The study also reported data from another evaluation of the program by the Texas Education Agency (Mitchell, 1987), which surveyed principals, finding, according to Lutz & Hutton, that:

---

appendix, and one (Bliss, 1992) can be found there under (Bliss, 1995).

[4] Another study by the California Commission on Teacher Credentialing found the attrition rates of Los Angeles Teacher Trainees who dropped out before they entered teaching, found in a report by the California Commission on Teacher Credentialing to be extremely high. Of the first cohort, 80.3% completed the first year of training and only 64.6% completed the second year and received a clear credential the year after (Wright, McKibbon, and Walton, 1987).

> The principals rated the beginning teachers as more knowledgeable than the AC interns on the eight program variables: reading, discipline management, classroom organization, planning, essential elements, ESL methodology, instructional techniques, and instructional models. The ratings of the AC interns on nine other areas of knowledge typically included in teacher preparation programs were slightly below average in seven areas compared with those of beginning teachers. It might therefore be assumed that pre-service teacher education programs are doing something right! (p. 250).

In the paragraph cited above, Lutz and Hutton wax enthusiastic about preservice teacher education programs that seemed in these data to outperform the alternative route. Later they wax enthusiastic about the alternative route, given results from another survey of principals, most of whom felt that alternative credential candidates who made through the program were comparable to other beginning teachers. Although Walsh cites Lutz and Hutton's enthusiastic feelings about the AC program, she does not accurately report the actual data from the study.

Walsh repeats this mistake in the appendix when she critiques a review of alternate certification programs (Darling-Hammond, 1992). She states that, "Darling-Hammond cites the findings from many studies that looked at alternative programs; but she does not include findings that show alternatively trained teachers are *at least as effective at raising academic achievement* as those who graduate from traditional programs," (p. A-3), citing Lutz and Hutton (1989), despite the fact that their study presented no empirical data on academic achievement of students and presented mixed evidence about the rated performance and retention rates of these recruits.

Two other studies Walsh cites do include student achievement data, but they do not, as she states, compare certified with uncertified teachers. Both deal with alternatively certified teachers who receive a substantial amount of education coursework while they are undertaking mentored teaching supervised by both university supervisors and classroom mentors.

4. **Miller, McKenna, & McKenna (1998)** is a well-designed quasi-experimental study of what the study's authors call a "carefully constructed" university-based alternate route

program for middle school teachers. Reflecting the characteristics of alternative routes endorsed by the National Commission on Teaching (1996), this program offered 15 to 25 credit hours of coursework before interns entered classrooms where they were intensively supervised and assisted by both university supervisors and school-based mentors while they completed additional coursework needed to meet full standard state certification requirements. [The 15 to 25 credit hours of pre-service coursework, only a portion of that required in this alternative route, is almost as extensive as the total requirements of Maryland's standard certificate, which requires 27 credits of education coursework (Walsh, p. 3).] Forty-one of these teachers were compared to a group of traditionally certified teachers matched for years of experience. Although the sample size is too small to meet Walsh's criteria (a point she seems to have forgotten here), and data are not provided on student pre-test scores, the study appears reasonably well conducted.

The traditionally trained teachers in this study felt somewhat more confident in their practice and scored slightly higher on the two sub-scales of an observation instrument used by trained observers to rate their teaching. However, these differences were not significant, and the authors report, without including the actual data analyses, that there were no significant differences in the two groups' student achievement by the 3rd year of practice after both had completed all of their education coursework. (The authors did not control for prior achievement levels of students; however, they stated that the initial differences in student achievement across groups were not significant.)

Because the design of this program was so different from many quick-entry alternative routes, Miller, McKenna, and McKenna note that their studies "provide no solace for those who believe that anyone with a bachelor's degree can be placed in a classroom and expect to be equally successful as those having completed traditional education programs.... The three studies

reported here support carefully constructed AC programs with extensive mentoring components, post-graduation training, regular in-service classes, and ongoing university supervision" (p. 174). This finding does not support Walsh's contentions throughout her paper that only general intelligence and subject matter knowledge make a difference for teacher effectiveness, or her claim that there is no support for teacher education and certification.

5. The other study on alternative certification cited by Walsh (**Bradshaw & Hawk, 1996**) was not published as a peer-reviewed article or research report – one of Walsh's criteria for rejecting the results of other reports. It was not retrieved in the brief time available to prepare this reply; however, it found, according to Walsh's separately-published appendix, "mixed, inconclusive results" regarding student achievement. No other analysis of the paper is offered.

In addition to its other inaccuracies, Walsh's review confuses alternative certification – a strategy that provides candidates with preparation that is differently packaged from whatever states deem "traditional" training -- with lack of certification – which generally indicates a lack of preparation. Having already missed this critical distinction, Walsh does not begin to attempt to sort out the effects of the differences in preparation experiences and outcomes associated with different models of teacher education. Thus, she does not note that program designs that include a comprehensive and coherent program of coursework and intensive mentoring (e.g. Miller, McKenna, & McKenna, 1996) have been found to produce more positive evaluations of candidate performance than models that forego most of this coursework and supervised support.

For example, a comparative study of more than 200 alternative certification candidates in New Hampshire, who are certified via three years of on-the-job training in lieu of formal preparation, found they were rated by their principals significantly lower than university-prepared teachers on instructional skills and instructional planning, and they rated their own

preparation significantly lower than did the university candidates (Jelmberg, 1995). To understand the outcomes of different approaches, studies of alternatives need to acknowledge the differences in program models.

Finally, Walsh cites two additional studies that include uncertified teachers, but she gets the findings wrong. Neither study shows that uncertified teachers do as well as certified teachers. One shows that the reverse is true.

6. In one study (**Goldhaber & Brewer, 2000**), the authors found that high school students who had a certified teacher in mathematics did significantly ***better***, after controlling for initial achievement and student demographic factors, than those who had uncertified teachers. The same trends were true in science, but the influences were somewhat smaller. In this sample, students of a small number of science teachers who held emergency or temporary certification (24 out of the 3,469 teachers in the overall sample) did no worse than the students of certified teachers, although they, too, did better than the students of uncertified teachers. Another analysis of these data (Darling-Hammond, Berry, & Thoreson, 2001) showed that in this sample most of the teachers on temporary / emergency certificates were experienced and most had education training comparable to that of the certified teachers. Most appeared to be already licensed teachers from out-of-state who were in the transition period to securing a new state license or experienced teachers teaching out of their main field. Only a third were new entrants whose characteristics may have suggested a content background with little education training. The students of this sub-sample of teachers had lower achievement gains in an analysis of co-variance that controlled for pre-test scores, content degrees, and experience than those of the more experienced and traditionally trained teachers.

7. Finally, Walsh cites a recently released study of Teach for America (TFA) by **Raymond et al. (2001)** that has not been published in a peer-reviewed journal. This study is relevant to

Walsh's discussion of the Resident Teacher Program through which she notes that many TFA recruits enter teaching in Maryland. However, the study did not compare certified to uncertified teachers, as Walsh claims. Although they had the data to do so, the authors chose not to examine how TFA teachers performed in comparison to trained or certified teachers. The study examined the influences of TFA teachers on student achievement scores, using regression methods that controlled for teacher experience and school demographics; thus, the comparison was between TFA recruits and other inexperienced teachers in high-minority schools in Houston -- where most underqualified teachers are placed. Since about 50% of Houston's new hires are uncertified and about 35% lacked a bachelors degree in the most recent year of the study, TFA recruits were compared to an extraordinarily underprepared set of teachers. In this comparison, students of TFA teachers did about as well as those of other inexperienced, largely untrained teachers, many of them without bachelors degrees. (Reviewers of this report have noted that the report should have compared TFA recruits to other BA holders and to prepared or certified teachers; based on the statistics shown, it is not clear whether the results of these comparisons would be favorable to TFA.)[5] The Raymond et al. report also indicated that minority students in Houston, who are disproportionately taught by these underprepared teachers, lose ground academically each year. In addition, fewer than 50% of African American and Latino 9th graders in Houston graduate from high school four years later (Haney, 2000; NCES, 2000). It would be hard to argue that the assignment of so many underprepared teachers to these students has nothing to do with these outcomes.

The report found that students of experienced teachers performed significantly better than students of inexperienced teachers, including TFA teachers. This, along with the report's finding

[5] Personal communications with Susanna Loeb, economics professor at Stanford University, and statistician William Billet.

that, over a three year period, between 60% and 100% of TFA candidates had left after their second year of teaching, raises questions about Teach for America's contribution to the education of Houston students. Earlier data from the Maryland Department of Education showed that TFA recruits in Baltimore had similar attrition rates, with 62 % gone by the third year of teaching (Darling-Hammond, 2000b).

These high attrition rates resemble those found in some other studies of short-term alternative routes (Darling-Hammond, 2000c) and suggest another important outcome of teacher preparation policies. Both the Houston study and Walsh's own review indicate that experienced teachers are more effective than inexperienced teachers (Walsh, pp. 5-6). Other research indicates that those who complete 5-year teacher education programs enter and stay in teaching at much higher rates than 4-year teacher education graduates, who stay in teaching at higher rates than teachers hired through alternatives offering only short-term summer training before full-time teaching (Andrew & Schwab, Darling-Hammond, 2000b). A recent NCES report notes that 29% of new teachers who had not had student teaching left teaching within five years – an entry strategy that is typical of emergency hires and many of the shorter term alternative routes – as compared to only 15% who had had student teaching (NCES, 2000, cited in Bolich, 2001).

Findings about the high attrition rates of those hired without full preparation for teaching raise questions about the cost-effectiveness of a recruitment strategy that relies on teachers with little preparation who are likely to leave the profession before they can learn to become effective with children. Meanwhile, the children they have taught – almost always the most disadvantaged students in the most disadvantaged schools – have not had the benefit a teacher with either professional knowledge or experience – two sources of greater teaching skill.

A recent study in Texas showed that teacher attrition costs school systems at least $8,000 for

each recruit who leaves in the first few years of teaching (Texas Center for Educational Research, 2000). It estimated that the high attrition of beginning teachers in Texas, who increasingly enter without preparation and often receive few supports in learning to teach, costs the state more than $200 million per year (p. 16). This and other studies of teacher attrition suggest that policymakers should consider both teaching effects and retention patterns when they think about how to recruit and prepare teachers.

Walsh chooses to ignore other studies showing that certified teachers do better than uncertified teachers.

8. One of these by **Hawk, Coble, & Swanson (1985), entitled "Certification: It Does Matter**," found – in contradiction to Walsh's statement cited above – that teachers' certification in mathematics has a large and statistically significant effect on student achievement gains in both general mathematics and, even more profoundly, in algebra. It compared pre- and post-test scores of students whose teachers who were certified in mathematics as compared to those of teachers with similar levels of experience who were uncertified in mathematics. This study is dismissed in one part of Walsh's review as too small (p. 34), so that its findings can be discounted with respect to certification. However, the size of the study does not appear to matter to Walsh when she chooses to cite it as a basis for arguing that only subject matter makes a difference to teaching effectiveness (p. 65). This double standard about the use of research permeates the report. A study is declared inadequate when it finds any contribution of teacher education or certification to any measure of teacher effectiveness but a study of comparable size or methodology – often the same study – is embraced elsewhere and used to support a different argument.

While the study does have a small sample size (it examined 36 teachers, paired by school, course, and ability level of students being taught), it is a reasonably well-controlled quasi-experimental design. The study does support the idea that subject matter knowledge matters to teaching. However, Walsh misrepresents the study as suggesting that only subject matter knowledge matters. The study did not directly examine the isolated effects of subject matter knowledge but the combined effects of subject matter knowledge and educational knowledge – including methods courses in the teaching of the content area – that are part of the certification requirements for an in-field credential. Authors Hawk, Coble, and Swanson concluded:

> The results of this study lend support to maintaining certification requirements as a mechanism to assure the public of qualified classroom teachers… " (p. 15).[6]

As this and other studies reviewed here suggest, and as I have noted elsewhere, content knowledge in combination with content pedagogical knowledge – that is, knowledge about how to teach the content – which, together with student teaching, constitute the major components of certification appear to make contributions to student learning that exceed the contributions of either component individually. An important policy point from this and other studies of certification is the fact that teachers would not have been guided or encouraged to acquire the content knowledge and content pedagogical knowledge represented by in-field certification unless there was a certification system in the state. While Walsh and the Fordham Foundation manifesto she endorses would turn all hiring decisions over to principals, it was principals in these schools – and in many others across the country – who hired and assigned out-of-field teachers to teach mathematics as well as other subjects. In a policy world that eliminates teacher

[6] As one of dozens of examples of general sloppiness, neither the Goldhaber and Brewer study nor the Hawk, Coble, and Swanson study cited by Walsh for this proposition even treated the question of whether "the most distinct problem in schools serving poor children is the number of teachers who are teaching subjects in which they have no expertise." Neither study examined or reported on the socioeconomic status of students or the distribution of teachers in schools serving different children.

BEST COPY AVAILABLE

certification, there would be no barrier to that practice occurring on an even more widespread basis.

9. Another, much larger study resulted in similar findings about teacher certification in California. **Fetler (1999)**[7] examined the relationship between school scores on the state's mathematics test and teachers' average experience levels and certification status in 795 high schools, after controlling for student poverty rates and test participation rates. It found that the percent of teachers on emergency credentials exerted a strong and highly significant negative influence on student achievement. The author concluded that, "After factoring out the effects of poverty, teacher experience and preparation are significantly related to achievement" (p. 13).

This study is cited but never discussed in Walsh's revised report. In her original appendix, Walsh applauded the study's methods but then sought to dismiss its findings with two inaccurate assertions. First, she suggested, incorrectly, that the study's results pertained to subject matter knowledge alone, not to the combination of subject matter and teaching knowledge represented by certification. She misread both the study and the requirements of California's credentialing system to make this claim, appearing to believe that individuals who have passed only the subject matter requirement of a content test are granted full credentials in California (they are not), that individuals who are certified through internship programs (California's alternative route) do not have to complete pedagogical requirements, and that individuals are hired on emergency permits solely if they lack content knowledge.[8] Walsh also

---

[7] High school characteristics and mathematics test results. Education Policy Analysis Archives, 7 (9): http://epaa.asu.edu/epaa/v7n9.html.

[8] As the study clearly states, California uses emergency permits for those who lack either subject matter competence or pedagogy or both. The requirement for a clear credential is passage of both subject matter competence and a set of pedagogical requirements, whether these are completed in a "traditional" or an "alternative" program, which in California would be an internship model requiring the candidates to meet the same standards as traditional programs. In fact, the composition of the emergency permit pool in California is nearly the opposite of what Walsh seems to surmise. This pool includes many teachers who have passed the subject matter test (or alternative content course requirements) in mathematics but who have not completed teacher education requirements. It also includes many

BEST COPY AVAILABLE

suggested, incorrectly, that the study "may have some basic methodology problems, by reaching conclusions using aggregated state-wide data." However, all of the study's data are aggregated to the school level, not the state level. (See the author's confirmation of this statement, below.)

In the original appendix,[9] Walsh stated:

> The article would be only be of interest if someone tried to assert that a teacher who knows no math could be a good math teacher. Any attempt to use this study as evidence against the practice of hiring alternatively trained teachers, as appears to be Darling-Hammond's implies (sic) and as Wilson et al. interpret it, loses all of its impact after reading Fetler.... In fact the author.... is primarily advocating ensuring that math teachers take more subject matter coursework, and is clearly disinterested in any effect that may be had from coursework in 'professional knowledge.'

The author, Mark Fetler, took strong issue with this interpretation of his findings or his intent.

When I shared Walsh's statement with Fetler, he wrote in reply:

> I am surprised that Kate Walsh makes those statements. I had a brief telephone conversation with her, but she was not forthcoming about her intent. Meeting the subject matter requirement involves both knowing the topic, e.g., Algebra, and the specific procedures needed to teach it in the classroom. Someone who knows how to solve quadratic equations, but does not know how to convey that information to children in a classroom, is a poor teacher. Both math subject knowledge and math pedagogy are essential. I believe that my study is consistent with these statements.... I would be surprised to hear of any research that demonstrated successful teaching that lacked either of those elements. My study supports the importance of appropriate credentials. Supposing that you could find people who know math to teach, if they lack the ability to communicate effectively with children, they will not succeed in the classroom and will create dissatisfied students, parents, colleagues, administrators, and board members. It will be a mess. Higher standards, not lower, are the solution.

Fetler also noted that, "the unit of analysis in my paper is the school. It is not based on statewide aggregated data."

---

teachers who have passed a basic skills test but have not completed either the subject matter or teacher education requirements for a clear credential. It includes very few individuals who have completed teacher education requirements but who have not completed subject matter requirements, since demonstration of subject matter competence is a prerequisite for entering the student teaching or internship portion of teacher education in California. Furthermore, experienced teachers who may be teaching math out of field would generally have been included in Fetler's data set as credentialed, since out of field teaching is not monitored by the state through the data set he used.

California's experience is a good example of what happens when pressures and supports for hiring credentialed teachers are relaxed. After nearly a decade of inadequate and unequal salaries, easy access to emergency permits and waivers, and few incentives for the training and equitable distribution of qualified teachers for high-need fields and locations, California, now one of the lowest-achieving states in the nation, found itself with more than 40,000 teachers teaching on emergency permits or waivers by 1999-2000. The vast majority of these teachers were teaching in a small number of urban school systems in schools with the highest proportions of low-income students and students of color. High-minority schools were nearly seven times as likely to have uncredentialed teachers as low-minority schools. Low-achieving schools were nearly five times as likely to have uncredentialed teachers as high-achieving schools [10] (SRI, 2000, pp. 41-43).

These results mirror those already noted in Baltimore, Houston, and other cities. The pattern appears across the country. For example, a recent series in the Chicago Sun Times[11] documented that "children in the state's lowest-scoring, highest-minority and highest-poverty schools were roughly five times more likely to have teachers who had flunked at least one certification test" and were least likely to have teachers who were "correctly certified." The burden should be on those who argue against efforts to ensure minimally qualified teachers for all students to prove that the confluence of race, poverty, and low achievement with the presence of untrained and uncertified teachers does not further disadvantage our nation's most vulnerable students.

---

[9] The original appendix was included in Walsh's draft dated July 23, 2001. Her final complete appendix modifies this statement only slightly, stating, "The author's principal and clear lament is the lack of subject matter knowledge in mathematics, with little mention at all of education coursework that may be lacking."

[10] High-minority schools were defined as those with more than 90% students of color; low-minority schools had fewer than 30%. High-achieving schools were defined as those in the top quartile of achievement on the SAT-9 tests used by the state; low-achieving schools were those in the bottom quartile.

[11] Rosalind Rossi, "Teacher woes worst in poor schools," Chicago Sun Times, October 10, 2001.

Evidence about Preservice Teacher Education

For the proposition that "there is little evidence that the content and skills taught in preservice education coursework is (sic) either retained or effective" (p. 7), Walsh cites two articles (Murnane, 1983; Veenman, 1984) from among the many dozens of studies of teacher education that could have been retrieved from the peer-reviewed literature, had she done a search. Both of these are very old pieces, published long before recent reforms in teacher education. Neither of them makes any statement in support of Walsh's claim.

1. **Veenman (1984)** describes the most frequently cited problems by novice teachers. These included concerns about topics ranging from classroom management to teaching loads and class sizes. Nowhere in the article does he suggest that what teachers learned in preservice education was not retained or effective. In fact, he notes that researchers should look more to the conditions of schooling than to teacher education for explanations for many of the problems beginning teachers cite. Veenman notes that the outcomes of teacher education may vary by characteristics of programs, citing studies finding that those who had had more intense student teaching, more competency-oriented teacher education coursework, or who were more satisfied with their teacher education experiences reported fewer problems in the classroom.

2. **Murnane's (1983)** article is not an empirical study but a brief commentary on the work of another author who proposed the development of doctoral degrees for teacher leaders. While he questions the value of doctoral education for developing pedagogical skills (as would I), Murnane is careful to point out that there are forms of teacher education that may be helpful, and that lack of evidence in large data sets about the effects of preservice education may be related to the lack of data collected on the topic at that time, nearly 20 years ago. (See additional discussion of this point under "Evidence about Verbal Ability" below.)

3. Walsh ignores the findings of other studies on this topic, including some she has cited for other propositions. She critiques Evertson, Hawley, and Zlotnik (1985) for their interpretation of the findings of Edward **Begle (1979),** "a respected mathematician" regarding his findings about teachers' subject matter preparation (p. 34). In one of the few early data sets providing evidence about teacher preparation -- a mammoth study of 112,000 students conducted through the National Longitudinal Study of Mathematical Abilities -- Begle (reported in Begle & Geeslin, 1972 and, with additional data, in Begle, 1979) found that measures of teacher subject matter knowledge did not exert strong influences on student achievement. He also found that coursework in mathematics methods had a stronger effect on student achievement than higher-level coursework in the subject matter (discussed in Begle, 1979). On the lack of influence of subject matter knowledge in his earlier study (Begle & Geeslin, 1972) Begle noted, and Walsh reports, that the teachers in the study may have had stronger content knowledge than the norm, since they had all been accepted to a National Science Foundation Summer Institute. This is an appropriate point.

However, Walsh chooses to ignore Begle's findings about the value of education coursework. She does not explain why. Walsh cites Begle's work at several points in her text, and refers readers to her appendix for a review of his work that is no longer there. In her separately-published appendix, Walsh admits of Begle (1979) that, "this is a scholarly work, employing defensible analyses at the time it was written for examining the data." She then nonetheless sought to dismiss it with a vague statement about possible aggregation bias (although achievement data were aggregated only to the classroom level), "too many variables" in the data set, and "much greater variance in the number of subject matter courses teachers took than the number of methodology courses they took." This last complaint is

particularly odd. The implications of greater variability in subject matter courses contradicts the point she makes above (concerning Begle & Geeslin, 1972), and would generally make it easier to find effects, if they are there to be found, rather than harder. In another instance (regarding Byrne, 1983), Walsh notes, correctly, that the limited variability in subject matter coursework levels may have made effects more difficult to find. Walsh seems confused about the research findings and their implications but clear about her goal of discrediting any results that support the value of teachers learning about how to teach their content to others.

4. Monk (1994) offers similar findings on this question from a more recent data set that incorporates more fine-grained variables about teacher education. Using data on 2,829 students from the Longitudinal Study of American Youth, Monk (1994) found that teachers' content preparation, as measured by coursework in the subject field, is positively related to student achievement in mathematics and science but that the relationship is curvilinear, with diminishing returns to student achievement of teachers' subject matter courses above a threshold level (e.g., five courses in mathematics). In addition, teacher education coursework (e.g. methods courses in the content area) had a positive effect on student learning and sometimes had "more powerful effects than additional preparation in the content area" (p. 142). Monk concluded that "a good grasp of one's subject area is a necessary but not a sufficient condition for effective teaching" (p. 142).

Monk told me that when Walsh first shared her brief appendix review of his work with him, he was surprised that she had used his work to emphasize the importance of subject matter knowledge without acknowledging his findings on the value of education courses. He noted in an email to me that he had communicated to Walsh that:

> My study of relationships between teacher course taking experiences and subsequent student gains in performance showed that the number of both content courses and

content-specific pedagogy courses in a teacher's background is positively related to pupil test score gains in the relevant content area. It is misleading to report the positive results for the content courses and to not acknowledge the positive results for the pedagogy courses.

After Monk communicated with Walsh, she did acknowledge in her appendix that Monk's study provides support for the contention that education coursework has a positive effect on teaching performance; however, she did not incorporate this admission in her claims that "not one" of the studies ever cited on this topic provides such support.

5. In addition to newer databases that allow some large-scale examinations of the influences of teacher education variables on student achievement, recent studies have begun to look at the outcomes of different teacher education program designs. For example, studies of 5-year teacher education programs—programs that include a bachelor's degree in the discipline plus an additional year of education study and extended student teaching—have found graduates to be more confident and better rated than graduates of 4-year programs in the same institutions and as effective as more senior teachers, as well as more likely to enter and remain in teaching **(Andrew & Schwab, 1995; Denton & Peters, 1988)**. Walsh does not review or cite any of these studies, even those that were available for her information from previous research she claims to have scrutinized.

The Influence of Verbal Ability on Teacher Effectiveness

There is little disagreement about the fact that verbal ability and subject matter knowledge influence teacher effectiveness, although Walsh tries to set up a straw man by suggesting, inaccurately, that some researchers, including myself, have argued otherwise. (See the section on "Misrepresentations of Research" below.) There are two areas of real disagreement, however. One is whether verbal ability alone is the only or best measure of teacher effectiveness. The other is how to evaluate the size of relative contributions of various

kinds of knowledge to teacher effectiveness. As examples cited earlier illustrate, the literature on teacher characteristics and their effects on teacher performance has been a captive of the measures most likely to be available in large data sets at any moment in time. While there are many studies evaluating the influences of teachers' standardized test scores, especially measures of verbal or general academic ability, because these variables have been readily available in large-scale data sets since the 1960s, data on teachers' course-taking backgrounds or teacher education experiences have only been included in large data sets only since the early 1990s. Thus, there are more studies finding influences of variables that have most often been measured.

1. Walsh uses an article by **Murnane (1983)** written nearly 20 years ago to argue for the primacy of verbal ability as a correlate of teacher effectiveness. She states, illogically, that, "to concede this relationship would mean acknowledging that formal teacher preparation is not as critical to student achievement as some would advocate" (p. 41). However, Murnane pointed out in his article that evidence about the influence of verbal ability was partly a function of the fact that standardized test scores were one of the few variables about teachers available in large-scale databases at that time, which did not include good measures of teacher education. In discussing the results on verbal ability, he diverges from Walsh's interpretation:

> Clearly one should not interpret these results as indicating that intellectual ability should be the sole criterion used in recruiting teachers or that formal teacher training cannot make a difference. In fact, the lack of evidence supporting formal preservice training as a source of competence may be to some extent a result of limitations in the available data. For example, all databases suitable for examining the correlates of teaching effectiveness as measured by student achievement gains pertain to a single school district. Since there is less variation in training among teachers within a district than among teachers in the country at large, these databases do not permit the most powerful possible tests of the efficacy of alternative teacher training programs (p. 565).

2. Walsh tries to use another article by **Greenwald, Hedges, and Laine (1996)** as evidence that verbal ability is the only critical variable influencing teacher effectiveness, and

misrepresents a communication she had with Larry Hedges, one of the study's authors, regarding the appropriate interpretation of his findings. Characterizing Greenwald, Hedges, and Laine's article as "a sound review of 60 studies," she then criticizes a direct reference to its findings in a report by the National Commission on Teaching and America's Future (Walsh, p. 17). Her criticism first alludes, incorrectly, to a chart in the Commission's report (which in fact referred to another study,[12]) then she criticizes the interpretation of the chart. The correct chart in the Commission's report (figure 5, entitled "Effects of Educational Investments" in Darling-Hammond, 1997, p. 9) was reproduced directly from Greenwald, Hedges, and Laine's table 7, column 1 (p. 379) with the same variable labels and statistics as presented in the original source. It describes the size of increase in student achievement for every $500 spent on several different kinds of investments. Here is a reproduction of the table from Greenwald et al.'s study:

Table 7 – The effect of $500* per student on achievement**

	Sample	
Input Variable	**Full Analysis**	**Publication bias robustness**
Per pupil expenditure	0.15	0.15
Teacher education	0.22	0.20
Teacher experience	0.18	0.17
Teacher salary	0.16	0.08
Teacher/pupil ratio	0.04	0.04

*1993-94 dollars

**All achievement outcomes are in standard deviation units.

In explaining the table, the study's authors noted that:

> The magnitudes (of the effects) for teacher education and teacher experience are higher than, but of the same magnitude, as PPE (per pupil expenditures). That is, one would expect comparable and substantial increases in achievement if resources were targeted to selecting (or retaining) more educated or more experienced teachers (p. 380).

[12] Walsh states that, "L. Darling-Hammond … presents a chart using an ambiguous term 'Teacher Qualifications' which accounted for nearly half of the student achievement gains." (p. 17). The chart to which Walsh alludes actually referred to another study by Ferguson (1996) and was clearly labeled as such. Another chart next to this one was drawn directly from a table in the Greenwald, Hedges, and Laine study, and was also clearly marked.

The Commission used this finding, as Greenwald, Hedges, and Laine had done, as an indicator that investments in teacher education showed stronger influences on pupil achievement gains than investments in other resources, like reduced teacher/pupil ratios. We noted in discussing their overall study that they had found evidence of the influences of teacher ability and experience, along with teacher education. However, Walsh criticizes the Commission's two-sentence characterization of the research (which she calls a discussion "in considerable detail") for failing to note that Greenwald, Hedges, and Laine found more studies supporting the influences of teacher verbal ability on achievement than what they labeled "teacher education" (measured in their study as masters degrees because this was the most widely used measure in large data sets.) She suggests that Hedges disagrees with the Commission's characterization, a view that Hedges clarified was inaccurate when I spoke to him. He indicated that Walsh had not revealed her intent or her interpretation of his findings when she contacted him, and wrote the following to explain his own view of the proper interpretation of his findings:

> It is true that the relationship between teacher verbal ability and student achievement is relatively large and consistent across the few studies that have examined it. However this does not imply that investing in teacher ability (among possibly poorly qualified teachers) is a cost-effective way to enhance student achievement. There are two reasons. First, teacher ability (among qualified teachers) may be more expensive than other resources that could be purchased to improve achievement. That is, there could be a strong relationship but high cost. Second, and more important, the relations found in the studies Greenwald, Hedges, and Laine (1996) reviewed were studies of practicing teachers. There is no reason to expect that the same relation holds among those who are not part of the teaching workforce.

The point here, similar to that made by Murnane (above), is not that verbal ability is not important, but that the evidence does not prove it is the only important contributor or the most efficient way to achieve teacher effectiveness. In fact, what most current certification systems seek to do as they combine tests of basic skills and general academic ability, subject matter, and teaching knowledge with evidence of successful supervised clinical experience and coursework

focused on specific teaching knowledge and skills is to help candidates assemble the several sources of support for their expertise in a more coherent way than would otherwise be the case.

In pursuit of her argument that only verbal ability makes a difference, Walsh seeks to discount other studies that have found strong influences of teacher certification test scores on teacher effectiveness as being relevant only to the measurement of verbal ability and irrelevant to the broader question of teacher certification. These studies are also misrepresented.

3. In her discussion of **Schalock (1979)** in the appendix (B13), Walsh seeks to dismiss his review's findings about the limited evidence regarding the relationships between teachers' measured intelligence and other indicators of effectiveness because the review is "old, old!!" and because, she argues,

> More recent research such as Summers and Wolfe, 1977; Ferguson, 1991; Ferguson & Womack, 1996 *(sic)*; Murnane, 1983; Hanushek, 1971; Strauss and Sawyer, 1986 suggest that intelligence (measured by SAT, verbal ability tests and college selectivity) are indeed substantially important.

Aside from the facts that two of these "more recent" studies pre-date the review she dismisses as "old, old!" and one (Murnane, 1983) is not a study at all, Walsh here cites two studies that she dismisses elsewhere for "aggregation bias" (Ferguson, 1991 and Strauss & Sawyer, 1986, see Walsh, p. 27) and another (Ferguson & Womack, 1993) that she dismisses without stating a reason (see discussion of Wilson et al., in Appendix B). The reader is referred to Appendix B for reviews of these issues, but the studies are not included there.

3. Walsh cites **Ferguson (1991)** for a number of her propositions, including the fact that teacher quality matters (p. 5), that teacher race does not matter (p. 6), and that verbal ability matters (p. 6). Later, she claims – when she wants to dismiss the study for its findings about teacher education and certification – that the study suffers from aggregation bias , a concern I address in the next section on methodological issues. Ferguson's study found, in an analysis of

nearly 900 Texas school districts that controlled for student background and district characteristics, that combined measures of teachers' expertise—scores on a state teacher licensing examination, master's degrees, and experience—accounted for more of the inter-district variation in students' reading and mathematics achievement (and achievement gains) in grades 1 through 11 than student socioeconomic status. An additional, smaller contribution to student achievement was made by lower pupil-teacher ratios and smaller schools in the elementary grades. The effects were so strong, and the variations in teacher expertise so great, that after controlling for socioeconomic status, the large disparities in achievement between black and white students were almost entirely accounted for by differences in the qualifications of their teachers.

As I noted in an earlier review of this study (Darling-Hammond, 2000), of the teacher qualifications variables, the strongest relationship was found for scores on the TECAT, a state licensing examination described by the test developer as a test that measures basic skills and professional knowledge. The Texas Education Agency's published outline of the test content shows that it seeks to measure verbal ability, logical thinking, research skills, and a set of items on professional knowledge. Walsh takes issue with this description of the test and argues that the study does not support the value of teacher certification because the test should be considered primarily a basic literacy test. In Walsh's view, this makes it irrelevant to the question of teacher preparation. She also argues that the relatively smaller influence of master's degrees in Ferguson's study (which accounted for about 5% of the explained variance) means that teacher education is unimportant, and she criticizes the fact that I discuss the three variables associated with teacher quality (TECAT scores, experience, and masters degrees) in combination, although this is also the way in which Ferguson discusses them at several points in his analysis.

Walsh's arguments are illogical in several ways. First, while it is true the TECAT measures basic skills, it is also measures other academic abilities and professional knowledge, as confirmed by the test maker's documentation and administering agency's descriptions. We have no way of knowing the relative importance of its various components and no basis for making judgments contrary to the claims of the developers. In addition, the test would not exist at all if there were not a state certification system requiring it. Like all of the other variables one can evaluate in studies of this kind, the test scores are a rough proxy for many aspects of teacher capacity that may matter for their performance. In a regression equation of this sort where one variable stands in for others for which data are not available, it undoubtedly captures the effects of other unmeasured factors. Even if it were true that the test was a weak measure of professional knowledge – something we cannot know from these data – this would not mean that professional knowledge is unimportant or that verbal ability is the only important variable for predicting teaching ability. Other variables matter as well. Finally, as Hedges notes above, since the Ferguson study was based on practicing teachers, its findings do not shed light on the relative effectiveness of non-teachers who might score differently on the tests.

Masters degrees and experience are other very partial measures of teacher knowledge and skill that show a modest effect in this study and a larger effect in Ferguson and Ladd's (1996) similar study in Alabama that included a weaker test measure of pre-college general skills (the ACT). However, masters degrees are also a very crude proxy for teacher education, given the wide variability in the content of masters degrees pursued by teachers, many of which have been pointed at jobs outside of teaching, such as administration, counseling, measurement and evaluation, and the like. Thus, there is reason to expect that some masters degree studies would affect teaching ability, but not much reason to expect the effect of masters degrees as an

undifferentiated variable to be uniform or large in the aggregate, a point I have made in earlier commentary (Darling-Hammond, 2000a), and one made by Goldhaber and Brewer (1998, 2000), whose research has noted the greater influence of both bachelors and masters degrees in the content area taught (e.g. mathematics or mathematics education) as compared to undifferentiated degrees.

It makes more sense to consider these variables together as proxies for expertise than to treat them as mythically precise measures of totally unrelated constructs. As I have argued elsewhere, research on teaching suggests a view of expertise that includes general knowledge and ability, verbal ability, and subject matter knowledge as foundations; abilities to plan, organize, and implement complex tasks as additional factors; knowledge of teaching, learning, and children as critical for translating ideas into useful learning experiences; and experience as a basis for aggregating and applying knowledge in non-routine situations (Darling-Hammond, 2000a). David Berliner's studies of expertise in teaching, for example, include experience along with several other traits as a critical aspect of expertise (see e.g. Berliner, 1986). All of these factors combine to make teachers effective; furthermore, one cannot fully partial out the effects of one factor as opposed to another as many are highly correlated.

5. Walsh also cites **Strauss and Sawyer (1986)** for her proposition that verbal ability matters (p. 6), but fails to report the study's actual findings. In a study of 145 school districts in North Carolina, these researchers found that teachers' average scores on the National Teacher Examinations (NTE) had a strong influence on average school district test performance. Although the authors did not specify which portion(s) of the NTE were used as measures, the Weighted Common Examinations Test (WCET) was required in North Carolina at that time The WCET included separate subtests measuring general knowledge and professional knowledge

about teaching. Walsh apparently wants to count this as a test of verbal ability, but does not acknowledge the Professional Knowledge Examination portion of the test.

The authors found that, taking into account per-capita income, student race, district capital assets, student plans to attend college, and pupil/teacher ratios, teachers' certification test scores had a strikingly large effect on students' failure rates on the state competency examinations: a 1% increase in teacher quality (as measured by NTE scores) was associated with a 3 to 5% decline in the percentage of students failing the exam. The authors' conclusion is similar to Ferguson's (1991):

> Of the inputs which are potentially policy-controllable (teacher quality, teacher numbers via the pupil-teacher ratio and capital stock), our analysis indicates quite clearly that improving the quality of teachers in the classroom will do more for students who are most educationally at risk, those prone to fail, than reducing the class size or improving the capital stock by any reasonable margin which would be available to policy makers (p. 47).

The same illogic holds in regards to the dismissal of this study as the previous one.

In addition to questions about the content of tests used in various studies, the measures that appear in large data sets are always relatively crude proxies for the constructs under study, so it is impossible to know with great precision exactly what trait is being represented when a variable shows an effect. For example, scores on tests of academic ability like the SAT have generally been strongly correlated with scores on ETS subject matter and professional knowledge tests (Gitomer, Latham, and Zimek, 1999); in eras when higher degrees were less common (e.g. pre-1980), verbal ability scores were also strongly correlated with masters degrees. Where certification tests are in place, test scores correlate with certification status. And both certification status and masters degrees typically correlate with teacher experience, since most states require teachers to obtain certification in order to remain in the workforce and most teachers have traditionally secured masters degrees by

taking courses over time while teaching. (This is changing to some extent where beginning teachers are being trained in post-baccalaureate or 5-year programs and sometimes enter the workforce with a masters degree).

These interrelationships do not invalidate studies that have used one or more of these variables, but they are one reason why it is difficult to say with certainty which of these measures – or other unmeasured variables that are related to them – are associated with measured effects. The correlational studies that Walsh relies on almost exclusively do not establish causation; they point to possible relationships for further, more fine-grained exploration. However, Walsh often dismisses both the larger studies and the more fine-grained studies from consideration, at least when the findings do not suit her predilections.

6. Walsh also cites **Ferguson & Womack (1993)** her proposition that verbal ability matters most, although the reason for this is unclear. This study of more than 250 candidates from a single teacher education program examined the influences on 13 dimensions of teaching performance of education and subject matter coursework, NTE subject matter test scores, and GPA in the student's major. The ratings of performance were based on detailed descriptors of teaching on 107 items evaluated by subject matter specialists and education supervisors. The authors found that the amount of education coursework completed by teachers explained more than four times the variance in teacher performance than did measures of content knowledge (NTE specialty scores and GPA in the major). It is possible that Walsh cites this study as support for verbal ability influences because she has confused the NTE specialty tests of subject matter knowledge with other components of the NTE battery measuring general academic ability. In any event, the strength of the relationship was very small. Given her willingness to cite the study for a very weak finding about verbal

ability, it is interesting that she does not cite it for its much stronger finding that education coursework mattered for teaching performance.

In her separately-published appendix, Walsh seeks to dismiss the Ferguson & Womack study because it is limited to a single institution[13] and uses "supervisor's evaluations" as the measure of performance. As noted earlier, she is willing to use studies based on such measures for her own claims, despite her assertions that they should not be included. More important, in this study the ratings are not the global ratings from school principals that have often been found to be relatively low in reliability. They are lower-inference ratings based on a detailed protocol used by subject matter specialists and university supervisors, which are typically more reliable. In addition, the limitations on generalizability created by the use of a single institution are not fatal to consideration of the findings. They require that the study be considered in the context of other studies on similar questions using different samples. Such studies have been conducted.

7. In a similar study which compared relative influences of different kinds of knowledge on 12 dimensions of teacher performance for more than 270 teachers, **Guyton and Farokhi (1987)** found consistent strong, positive relationships between teacher education coursework performance and teacher performance in the classroom as measured through a standardized observation instrument (the Georgia Teacher Performance Assessment Instrument), while relationships between classroom performance and subject matter test scores were positive but insignificant and relationships between classroom performance and basic academic skill scores were almost nonexistent. (The two measures of basic academic skills were the Georgia Regents'

[13] One odd criticism is that the institution, Arkansas Tech, has "low entrance requirements, making it unlikely that enough variance in student ability, background and coursework is present to reflect a broader population. The variance may be too narrow or at least skewed." Walsh seems to be unaware that the variance in student ability measures is usually much larger in large state universities like this one than it is in more selective colleges, thus making some kinds of inferences more, rather than less supportable. The more appropriate question about single institution studies is whether they may generalize to unlike institutions, a legitimate point that Walsh does not raise, and that should be answered by conducting studies within and across institutional contexts.

test, a required examination for public university students, for which the researchers used reading and essay scores, and the states' Teacher Competency Test.)

The researchers noted that extensive reliability studies had been conducted to support the reliability of the TPAI performance measure, which was used statewide as an assessment for certification. Walsh eliminates this study from consideration because it is a single institution study and refers the reader to Appendix B for her review (p. 25). In her appendix Walsh further critiques the study with a note about the unreliability of supervisors' ratings, again failing to distinguish the research on principals' general teacher evaluation ratings from the research on the reliability of the TPAI as an observational instrument. She also apparently failed to read the study carefully, voicing questions about why the numbers of teachers in different comparisons differ, not having noted the authors' explanation that all correlations depended upon the number of teachers for whom data on both variables was available and that they gave sample sizes for the different variables (p. B11).

Whereas Walsh tries to paint an unambiguous picture about the value of such measures as verbal ability (suggesting, for example, that these scores be reported statewide as a primary measure of accountability) and the lack of value of teacher education, the real picture is decidedly more complex. Her evidence for this claim confuses measures of verbal ability with measures of professional knowledge and subject matter knowledge, and often includes studies that actually show influences of these other kinds of knowledge that are at least as strong as measures of verbal ability. The world is just not as simple as Walsh would like to make it appear. Even strong advocates of the notion that academic ability matters are not willing to make the kinds of over-assertions Walsh urges. For example, Hanushek (1992), whom Walsh cites repeatedly for her defense of verbal ability as a key measure concludes:

> The closest thing to a consistent finding among the studies is that "smarter" teachers who perform well on verbal ability tests do better in the classroom. Even for that the evidence is not very strong (p. 116).

While it would be ridiculous to argue that verbal ability and subject matter knowledge do not matter for teaching, it is equally ridiculous to argue that knowledge of teaching and learning and the opportunity to learn to teach under the close supervision of a master teacher through student teaching and other guided experiences do not matter at all. The literature just does not support this reading or the policy implications that Walsh would draw.

The Academic Ability of Teachers who Lack Certification

Another argument made by those who would eliminate certification is that an unconstrained market would allow the recruitment of individuals with higher verbal or general academic ability that do not now enter teaching. While it is probable that some individuals would choose to teach if they did not have to prepare, it is not clear that most would be more academically able, that they would be better teachers, or that they would stay long in teaching. It is also unlikely that given current wages, individuals who are now preparing for much higher-paying careers in medicine, the law, engineering, and other professions that require much more onerous preparation and licensing processes would choose to teach simply because they did not have to be certified.

Labor market contexts are relevant to this question. The qualifications of individuals preparing for teaching improved noticeably between the early 1980s and the early 1990s in terms of both academic attainment and ability measures, in part because of the changes in admissions requirements to teacher education adopted by states and universities but also likely because of the substantial increases in real wages for teachers that occurred during the 1980s. Whereas prospective teachers were disproportionately drawn from the bottom quartile of college students

in the early 1980s (Lanier & Little, 1986), both grades and test scores improved for teacher candidates by the 1990s.

The Recent College Graduates Survey, which tracks college graduates into the labor market, found that the grade point averages of newly qualified teachers in 1990 were higher than those of the average college graduate, with 51% earning a GPA of 3.25 or better as compared to 40% of all graduates (Gray et al., 1993). However, average GPAs were significantly ***lower*** for the 15% of college graduates entering teaching who were neither certified nor eligible for certification than for those who had prepared to teach. Most of the uncertified entrants (57%) had grade point averages below 3.25, and 20% had GPAs below 2.25. Attrition was also high for the untrained candidates. By the time of the survey (one year later), only one-third of the uncertified entrants were still engaged in teaching as their primary jobs (Gray et al., 1993).

In addition, the Educational Testing Service found that among 270,000 test-takers in 1995 through 1997, college admissions test scores were highly correlated with initial teacher licensing scores (Praxis I and Praxis II), and the lowest average scores on both kinds of tests were those held by individuals who entered teaching without preparation. (Walsh describes this 14% of the sample as an "error" in the study since the individuals had not enrolled in a teacher education program; she misunderstands the fact that these Praxis test-takers were the entrants to teaching who used emergency or alternative routes.[14]

Prepared teachers scored much higher. While students who prepare to enter fields other than teaching have higher average test scores on measures like the SAT than do those preparing to enter teaching, the differences are significant only for elementary and special education majors; there is not a significant difference for prospective secondary teachers earning a

[14] Some may also have been those teachers who needed to take the Praxis as an entrance examination for a teacher education program.

disciplinary degree along with their teaching certificate (Gitomer, Latham, and Zimek, 1999). Here, too, the narrowing of this gap between prospective teachers and others is likely a function of the more rigorous admissions requirements for teacher education enacted in most states and the growth in wages between the early 1980s and the mid-1990s.

Finally, the study found that graduates of NCATE-accredited colleges of education passed the Praxis subject matter tests for teacher licensing at a significantly higher rate than did graduates of unaccredited programs, boosting their chances of passing the examination by nearly 10 percent (Gitomer, Latham, and Zimek, 1999). Walsh suggests that this higher Praxis pass rate might simply reflect the fact that NCATE schools could be located in states with low cutoff scores. However, additional analyses of the data by ETS and another independent study[15] indicate that this is not the case. A more likely explanation is that NCATE's requirements that colleges demonstrate how they screen applicants for general ability and that they ensure strong content backgrounds translate into somewhat greater attention to these matters in institutions that are accredited. These data suggest that standards may increase the general as well as specialized qualifications of prospective teachers. They do not suggest that removal of certification requirements brings higher ability individuals into teaching.

It is important to recognize that labor market incentives operate among individuals actually entering teaching. For example, several studies of alternative certification programs found that the academic records of recruits varied substantially by teaching field, with alternatively-certified candidates in high demand shortage fields, such as mathematics and science, having much

[15] The ETS re-analysis is soon to be published. An earlier analysis of the federal Baccalaureate and Beyond data base found that 1993 graduates of NCATE-accredited teacher education programs were about 50% more likely to have scored above the 50th percentile on SAT and ACT tests than graduates of non-NCATE teacher education programs (Shotel, 1998). NCATE graduates had also taken more social science, computer science, advanced foreign language credit, pre-college mathematics, and teaching coursework and fewer remedial English courses than non-NCATE graduates, with other areas being approximately equal (Shotel, 1998).

BEST COPY AVAILABLE

poorer academic records than candidates in other fields and than candidates from traditional teacher education programs in those same fields (see Natriello et al., 1990 re: New Jersey; Lutz and Hutton, 1989 re: Dallas; Stoddart, 1992 re: Los Angeles). It is unlikely that eliminating requirements for training would increase the career attractions to teaching for academically able candidates as much as increased wages would. Meanwhile, eliminating training requirements could result in a less well-qualified teaching force, especially if the elimination of certification standards also reduced pressures for competitive wages.

The Private School Argument

Finally, a claim sometimes made by opponents of teacher certification, including Walsh, is that private schools are more effective than public schools, and that this is because – or at least is not impeded by -- the fact that private school teachers are not certified. There are two major problems with the private school "proof": First, there are conflicting findings about the relative effectiveness of public and private schools, with credible evidence on both sides of the question. Second, most private school teachers are certified and an even larger majority have specific preparation for teaching, even when they have not sought to be credentialed.

On the effectiveness of private schools, Walsh cites Coleman, Hoffer, & Kilgore (1982), who examined data from the first wave of High school and Beyond surveys, conducted in 1980, and found evidence of higher performance for comparable students in Catholic and other private schools as compared to public schools. The researchers attributed their findings primarily to differences in student behavior across school sectors, measured by variables like lower rates of absenteeism, cutting class, and fighting, along with factors like more time spent on homework and higher individual student attendance. They also found that achievement was actually ***higher*** for comparable students who were in public schools that had these characteristics. Subsequent

studies have produced findings that favor both private and public schools after controlling for student characteristics and school organization (Bryk & Lee, 1992; Lee & Bryk, 1988; Lee, Dedrick, & Smith, 1991). Most studies have pointed to variables like school and class size, school organization, and curriculum differentiation as critical variables in determining both public and private school effectiveness. When these factors are controlled, public school students often do as well or better than private school students in schools with similar features.

Furthermore, differences in the preparation of public and private school personnel are not as large as many people assume. More than 30 states certify private school personnel (Feistritzer, 1984), and, when Coleman did his analysis, more than 85% of private and parochial school teachers were certified, as compared to about 95% of public school teachers (NCES, 1985). This has changed only slightly in the years since. Although certification is not required for private school teachers in all states, only 34% of private school teachers in 1993-94 (the most recent year for which national data are available), were not certified in their primary assignment field. Some of these teachers were certified in fields other than their primary assignment field. Many undertook teacher preparation, even though they did not apply maintain a state license or certificate. In 1993-94, public and private school teachers were almost equally likely to have received an undergraduate degree in education (68.9% for public vs. 61.5% for private elementary teachers and 19.8% for public vs. 19.3% for private secondary teachers) (NCES, 1997, p. 25). The education degree as an indicator of preparation is quite partial, since the education degree has waned as certification increasingly requires a content degree with an education minor or credential. The percentage of 1992-93 bachelor's degree recipients who had taken education courses was 87.1% for public school teachers and 71.6% for private school

teachers,[16] and the average number of education credits earned was 37.4 for public school teachers as compared to 35.2 for private school teachers (NCES, 1997, table A-51).[17]

Public school teachers were also somewhat more likely to have taken subject matter degrees in their teaching fields than private school teachers. For example, 66% of public school mathematics teachers held a major or minor in the field, as compared to 58% of those in private school. (Goldhaber and Brewer, 2000 reported a similar finding.) The same differentials hold in other fields to somewhat lesser extents. The greater content preparation of public school teachers is likely a function of the fact that certification has required increasing amounts of subject matter coursework in the field to be taught, thus leveraging stronger content preparation for public school teachers in states where private school teachers are not required to hold certification. Almost all states now require certified teachers to hold at least a minor in the field to be taught, and many require a major in the field.

Finally, even if it were true that untrained teachers were unusually effective in some private schools for students of comparable initial achievement levels – a point about which there is no published evidence – it would be a large leap of faith to assume that such teachers would be equally effective in schools where many students have much greater educational needs and students are not pre-selected for their academic ability, their positive school attendance and behavior, and their parents' income and interests in education. There are very large differences

---

[16] The proportions who had taken other kinds of liberal arts coursework also differed little. For example, the proportion of 1992-93 bachelor's degree recipients who had taken college coursework in mathematics at the level of calculus and above was 18.3% in public schools and 16.9% in private schools; science was 77.2% vs. 73.5% (table A-51).

[17] These statistics pertain to the youngest teachers in public and private schools: 1992-'93 bachelors degree recipients hired by 1993-94. These teachers are the least likely to be certified, even though they have taken education coursework at rates nearly as high as public school teachers. This suggests that many of these teachers may have prepared to teach but did not seek or secure state certification. In 1993-94, NCES reports that about 36% of private school teachers held no certificate in their primary assignment field (the data are not presented regarding their certification in another field other than the primary teaching assignment). The rates of non-certification ranged

in the populations of students attending public and private schools in the United States.[18] These differences have important implications for teachers' knowledge and skills. It is one thing for a teacher to offer information in whatever manner comes instinctively to students who are academically able, have learned to learn independently, and are well-supported at home by educated parents, tutors, and other supports for their learning. It is quite another thing to teach by the seat of the pants when students do not have these learning supports at home and may present a variety of language and learning differences. Being effective with students who need substantial support for their learning requires greater diagnostic ability and knowledge of how to present information and structure experiences in ways that help them become successful. To the extent that systematic knowledge about how to organize curriculum and reach students with special learning needs is valuable, it stands to reason that it is most needed in the schools that serve most students with these needs.

Other Misrepresentations of Research Findings

The remainder of the review continues the kind of misrepresentations documented above, appearing to rely on the belief that readers will read its accusations, but will not read or understand the research itself. Although she prepared a draft appendix with 192 studies that sought to critique many of the studies she dismisses (often inaccurately), Appendix B, to which the reader is repeatedly referenced for reviews, includes only 14 studies. Throughout the report, the reader is referenced to examine this appendix for critiques of studies that do not appear there. The selection of research included in the new published version of the report's appendix is very

---

from 27% for those with 20 or more years of teaching experience to 51% for those with 3 or fewer years of teaching experience (NCES, 1997, table A3.14a).

[18] For example, while most private school students (52%) attend schools that are less than 10% minority, only 31% of public school students do (NCES, Digest of Education Statistics, 1999, p. 71, table 60 and p. 119, table 99). African American and Latino students are at least 50% more likely to attend public than private schools. (NCES, 1997, Table A2.13). Most low-income students and students of color now attend public schools in urban public school districts.

strange. Many strong studies -- some of the key citations in the field -- are omitted, along with the often flawed rationales for dismissing them that now appear in a separately-published appendix. Some much less important and less well-designed studies are included, with the apparent goal of critiquing their size or designs as though they represented the dozens of studies not mentioned or excluded. Thus, the paper does not include information regarding most of the studies Walsh claims she has reviewed and does not provide evidence for her claim that, of all the studies cited in support of teacher education and certification, "none bear up to scrutiny."

Here are just a few additional examples of the absurdity of this claim.

1. **Goldhaber & Brewer (2000).** In a string of citations, Walsh lists a study by Goldhaber and Brewer (2000), for its finding that teachers with a degree in their subject matter are more effective than those without such degrees. This study fits all of Walsh's desiderata: It is large (using a data set that includes more than 3,000 teachers), recent, and published in a peer-reviewed journal. However, Walsh does not cite the authors' findings that certification status has an even greater influence on teachers' effectiveness than a degree in the subject area. Later, Walsh states, "...most research indicates that the most distinct problem in schools serving poor children is the number of teachers who are teaching subjects in which they have no expertise (Goldhaber & Brewer, 2000; ... Hawk, Coble, & Swanson, 1985). These studies do not show that certification status, as an isolated variable, has any significant effect on the achievement level of children who are poor or minority." (p. A6), thus misrepresenting the findings of both studies. In fact, Goldhaber and Brewer wrote:

> Turning to an examination of the effect of teacher certification, we find that the type (standard, emergency, etc.) of certification a teacher holds is an important determinant of student outcomes. In mathematics, we find the students of teachers who are either not certified in their subject (in these data we cannot distinguish between no certification and certification out of subject area) or hold a private school certification do less well than students whose teachers hold a standard, probationary, or emergency certification in

> math. Roughly speaking, having a teacher with a standard certification in mathematics rather than a private school certification or a certification out of subject results in at least a 1.3 point increase in the mathematics test. This is equivalent to about 10% of the standard deviation on the 12th grade test, a little more than the impact of having a teacher with a BA and MA in mathematics. Though the effects are not as strong in magnitude or statistical significance, the pattern of results in science mimics that in mathematics. Teachers who hold private school certification or are not certified in their subject area have a negative (though not statistically significant) impact on science test scores (p. 139).

The authors note that the effect size of "having a teacher with a standard certification in mathematics rather than a private school certification or a certification out of subject" is "a little more than the impact of having a teacher with a BA and MA in mathematics." Of course, the certification itself includes requirements for subject matter knowledge as well as for knowledge of teaching and learning. In fact, certified mathematics teachers were more likely to have a degree in the field than non-certified teachers. The fact that the study found a significant effect of certification status even after controlling for whether teachers had a degree in their field and after controlling for experience suggests that whatever is represented by the certification variable has an influence above and beyond the influence of content knowledge and classroom experience.

2. Druva & Anderson (1983). This meta-analysis of 65 studies examined relationships between science teacher characteristics and teaching behaviors, student achievement in science, or both, using meta-analytic techniques to translate results from a wide range of studies into Pearson correlation coefficients in order to compare them. It found that ratings of teaching effectiveness by principals and students were most strongly correlated with the number of education courses taken, followed by student teaching grades, and teaching experience. On a teacher "effectiveness" scale comprised of many teaching behaviors associated in process-product research with student achievement, both science training (examined in 28 studies) and

education coursework and performance (examined in 47 studies) were related to effectiveness, as were teacher attitudes, values, and temperament. Associations with cognitive and affective student outcome measures were found for both science training and, to a somewhat smaller extent, for education coursework and performance, based on 34 studies for each of these sets of variables. The authors concluded that:

> Student outcomes are positively associated with the preparation of the teacher, especially science training, but also preparation in education and academic work generally.... While the hiring official seeking a new science teacher certainly must look beyond information on the teacher characteristics considered in this study, information on some of these characteristics certainly is worthy of inclusion in the decision-making process.... In general, the hiring official would be well advised to employ teachers with thorough preparation in both professional education and the sciences being taught. There is a relationship between teacher preparation programs and what their graduates do as teachers (p. 477).

Walsh seeks to dismiss the results of this study in part by misreporting them. She states the study "did not show the benefit of education coursework on student achievement" (p. 19), and that education coursework was not significantly related to student outcomes, although significance statistics were not reported in the study. This assertion is not supported by the authors' reported findings that both science coursework and education coursework showed a relationship to teacher effectiveness as defined by student outcomes (in both cases, though to a greater extent for science coursework) as well as teaching behaviors and ratings (reported in the case of education coursework only).[19]

3. Darling-Hammond (2000). Walsh criticizes and misquotes a study that this author conducted, which examined both the literature on teacher characteristics and student achievement and conducted a regression analysis of state-level data from the National

[19] For mysterious reasons, Walsh also objects to the creation of a composite education variable that logically includes education coursework, student teaching performance, and education GPA, and she objects to the fact that many of the studies cited are dissertations.

Assessment of Educational Progress and the Schools and Staffing Surveys (Darling-Hammond, 2000). The study found that measures of teacher preparation and certification were by far the strongest correlates of student achievement in reading and mathematics, both before and after controlling for student poverty and language status, accounting for between 40 and 60 percent of the total variance in state scores. The conclusion discussed a number of potential reasons for these large effects:

> The strength of the "well-qualified teacher" variable may be partly due to the fact that it is a proxy for both strong disciplinary knowledge (a major in the field taught) and substantial knowledge of education (full certification). If the two kinds of knowledge are interdependent as suggested in much of the literature, it makes sense that this variable would be more powerful than either subject matter knowledge or teaching knowledge alone. It is also possible that this variable captures other features of the state policy environment including general investments in, and commitment to, education, as well as aspects of the regulatory system for education, such as the extent to which standards are rigorous and the extent to which they are enforced.... Finally, there may be unmeasured correlations between the extent to which states enact and enforce high standards for teachers and the extent to which they have enacted other policies that are supportive of public schools. Although it does not appear that teaching standards are strongly related to investments regarding class sizes or to overall education spending, it is possible that there are other factors influencing student achievement which generally co-exist with teacher quality and which were unmeasured in these estimates.

Walsh seeks to invalidate these findings by raising two complaints, one of which is inaccurate and the other of which is a matter of legitimate discussion in the field. She states, incorrectly, that, "Darling-Hammond did not control for class size differences among the states" (p. 26). State-level differences in average class size were in fact included in the analyses, and the variable had a very small, insignificant effect. Walsh also complains that the state-level analyses suffer from aggregation bias because they used average student test scores – a critique she also levels against other studies she cited approvingly for their findings in other parts of the paper (see e.g. Ferguson, 1991; Strauss & Sawyer, 1986; Coleman, 1966).[20] There are legitimate

[20] In Walsh's original appendix, this study is further critiqued because the reviewer was not clear on the meaning of the term "out-of-field" in the study when referencing elementary school teachers. The article defined the

debates in the field on this point, and I addressed this question in the study itself, as I do again below in the section on "Methodological Issues." For purposes of tracking broad policy trends at the state level, analyses of state level data offer one useful lens. This perspective was shared by the nine reviewers who recommended this paper's publication in a peer-reviewed journal and a peer-reviewed research report series.

Finally, the literature review contained in this study is repeatedly mischaracterized throughout Walsh's paper and her appendix as minimizing or ignoring the influences of verbal ability and subject matter preparation for teaching.

On the relationship between academic ability and teacher effectiveness, Walsh states:

> Darling-Hammond (1999, p. 6) claims there is "little or no relationship between teachers' measured intelligence and their students' achievement." She supports this statement with two studies by Soar, Medley and Cocker (sic) (1983) and Schalock (1979). These two studies simply recycle research from the 1940s and earlier, none of which is retrievable for scrutiny (p. 21).

Walsh misrepresents this analysis by quoting a portion of a sentence out of context and citing the reviews that summarized research on IQ tests as an example of the inappropriate use of older studies. Here is what I actually said:

> While studies as long ago as the 1940s have found positive correlations between teaching performance and measures of teachers' intelligence (usually measured by IQ) or general academic ability (Hellfritsch, 1945; LaDuke, 1945; Rostker, 1945; Skinner, 1947), most relationships are small and statistically insignificant. Two reviews of such studies concluded that there is little or no relationship between teachers' measured intelligence and their students' achievement (Schalock, 1979; Soar, Medley, & Coker, 1983). Explanations for the lack of strong relationship between measures of IQ and teacher effectiveness have included the lack of variability among teachers in this measure and its tenuous relationship to actual performance (Vernon, 1965; Murnane, 1985). However,

---

proportion of "well-qualified teachers" as the proportion holding state certification and the equivalent of a major (either an undergraduate major or masters degree) in the field taught. For elementary teachers, the equivalent of a major was defined an elementary education degree for generalists who teach multiple subjects to the same group of students or as degree in the field taught for elementary specialists (e.g. reading, mathematics or mathematics education, special education). The study defined "out-of-field" for elementary teachers in the same way it was defined for secondary teachers: holding less than a minor or the equivalent in the fields described above (elementary education in the case of generalists or the specialist field (e.g. reading or mathematics in the case of specialists).

> other studies have suggested that teachers' verbal ability is related to student achievement (e.g., Bowles & Levin, 1968; Coleman et al., 1966; Hanushek, 1971), and that this relationship may be differentially strong for teachers of different types of students (Summers & Wolfe, 1975). Verbal ability, it is hypothesized, may be a more sensitive measure of teachers' abilities to convey ideas in clear and convincing ways (Murnane, 1985)."

Walsh's attempt to distort the text misses two critical points: First, studies of the relationship between IQ and teaching effectiveness (which I noted had found positive though small relationships) were primarily conducted before the 1960s, because IQ tests came into question as measures of ability at that time and were no longer often available in large data sets thereafter. Measures of verbal ability became more popular and widely available in data sets in the 1960s and following, and showed somewhat stronger relationships with teacher outcomes, as I reported in my summary. The studies I cited include many of the same ones that Walsh cites for this proposition – a point she does not acknowledge as she tries to suggest, inaccurately, that I minimize the value of measures of academic ability for teachers.[21]

On the topic of subject matter knowledge, Walsh also suggests on numerous occasions that I seek to minimize the importance of teachers' knowledge of content. She offers my work as an example of her sweeping statement that "certification advocates … offer evidence that knowledge of subject matter has little effect on teaching performance" (p. 19). Here is what I actually said in my brief summary of the literature, offering a nuanced analysis that clearly acknowledges the importance of subject matter knowledge for teaching and interprets the mixed results of studies in terms of what teachers may need to know in order to teach different things.

---

[21] For some mysterious reason, Walsh also tries to make a point that I differentiate (wrongly in her view) between cognitive ability or IQ and verbal ability (see her footnote 14, p. 8), despite the fact that this is a standard distinction in the literature made by many of the analysts Walsh herself quotes for support of the importance of verbal ability measures. Few measurement experts would argue that IQ, as it was defined and measured in the 1940s and '50s, represents the same construct as verbal ability, as Walsh seems to be invested in proving.

Byrne (1983) summarized the results of thirty studies relating teachers' subject matter knowledge to student achievement. The teacher knowledge measures were either a subject knowledge test (standardized or researcher-constructed) or number of college courses taken within the subject area. The results of these studies were mixed, with 17 showing a positive relationship and 14 showing no relationship. However, many of the "no relationship" studies, Byrne noted, had so little variability in the teacher knowledge measure that insignificant findings were almost inevitable. Ashton and Crocker (1987) found only 5 of 14 studies they reviewed exhibited a positive relationship between measures of subject matter knowledge and teacher performance.

It may be that these results are mixed because subject matter knowledge is a positive influence up to some level of basic competence in the subject but is less important thereafter. For example, a controlled study of middle school mathematics teachers, matched by years of experience and school setting, found that students of fully certified mathematics teachers experienced significantly larger gains in achievement than those taught by teachers not certified in mathematics. The differences in student gains were greater for algebra classes than general mathematics (Hawk, Coble, & Swanson, 1985). However, Begle and Geeslin (1972) found in a review of mathematics teaching that the absolute number of course credits in mathematics was not linearly related to teacher performance.

It makes sense that knowledge of the material to be taught is essential to good teaching, but also that returns to subject matter expertise would grow smaller beyond some minimal essential level which exceeds the demands of the curriculum being taught. This interpretation is supported by Monk's (1994) more recent study of mathematics and science achievement. Using data on 2,829 students from the Longitudinal Study of American Youth, Monk (1994) found that teachers' content preparation, as measured by coursework in the subject field, is positively related to student achievement in mathematics and science but that the relationship is curvilinear, with diminishing returns to student achievement of teachers' subject matter courses above a threshold level (e.g., five courses in mathematics).

It may also be that the measure of subject matter knowledge makes a difference in the findings. Measures of course-taking in a subject area have more frequently been found to be related to teacher performance than have scores on tests of subject matter knowledge. This might be because tests necessarily capture a narrower slice of any domain. Furthermore, in the United States, most teacher tests have used multiple-choice measures that are not very useful for assessing teachers' ability to analyze and apply knowledge. More authentic measures may capture more of the influence of subject matter knowledge on student learning. For example, a test of French language teachers' speaking skill was found to have significant correlation to students' achievement in speaking and listening (Carroll, 1975).

It seems logical that teachers' abilities to handle the complex tasks of teaching for higher-level learning are likely to be associated, to varying extents, with each of the variables reviewed above: verbal ability, adaptability and creativity, subject matter knowledge,

> understanding of teaching and learning, specific teaching skills, and experience in the classroom, as well as interactions among these variables. In addition, considerations of fit between the teaching assignment and the teacher's knowledge and experience are likely to influence teachers' effectiveness (Little, 1999), as are conditions that support teachers' individual teaching and the additive effect of teaching across classrooms, such as class sizes and pupil loads, planning time, opportunities to plan and problem solve with colleagues, and curricular supports including appropriate materials and equipment (Darling-Hammond, 1997b).

Finally, Walsh suggests in several places that I have characterized the research as indicating a "negative relationship between student outcomes and the NTE subject matter tests" (p. 19). In fact, I stated that "Studies of teachers' scores on the subject matter tests of the National Teacher Examinations (NTE) have found ***no consistent*** relationship between this measure of subject matter knowledge and teacher performance as measured by student outcomes or supervisory ratings. Most studies show small, statistically insignificant relationships, both positive and negative (Andrews, Blackmon & Mackey, 1980; Ayers & Qualls, 1979; Haney, Madaus, & Kreitzer, 1986; Quirk, Witten, & Weinberg, 1973; Summers & Wolfe, 1975)."[22] Walsh misrepresents this statement numerous times.

Methodological Issues

One of the ways that Walsh seeks to make much of the research on teacher education disappear is by suggesting that it is inappropriate to cite studies that are older, smaller, use measures of performance other than student achievement scores, are aggregated at a level above the classroom or the school, or are published in venues other than peer-reviewed journals.

As noted above, Walsh uses a double standard in selecting research to reject when it finds evidence of the influence of teacher education on student learning and research to cite for her own purposes. While she discounts the findings of many dissertation studies and technical

[22] Walsh makes a hash of the research cited here on the relationship between teacher test scores and measures of teacher effectiveness, but that issue is not at the center of this debate and will not be reviewed here, since there is

reports because they were not published in peer-reviewed journals, in making her own claims, she cites at least 15 studies that were not published in peer-reviewed journals or technical report series and at least 20 that were published before 1980, including some that she elsewhere dismissed from consideration because she did not like specific findings. For findings she likes, she also cites several that use supervisory ratings as the only measures of teacher effectiveness and others that she later dismisses for aggregation bias. Sometimes she represents the studies' findings accurately; sometimes not. Many of the studies she cites for various propositions do not contain the findings for which they are cited – or, in several cases, any data on the question at all.

I would not argue, as Walsh does, that none of these studies have value as contributions to the literature. However, the double standard she applies in using studies of different eras, sizes, aggregation levels, dependent variables, and publication statuses perhaps proves the point that to evaluate the weight of evidence in a field it is often necessary to triangulate findings that used different methods, over different time periods, and at different levels of aggregation to see where there is an accrual of evidence over time and across methods. Of course it is important to do this with appropriate attention to the methodological strengths and weaknesses of various studies and lines of research. Unfortunately, Walsh often does this poorly, frequently appearing to misunderstand critical research design issues. Below, I discuss the issues of study size and design, level of aggregation, choice of dependent variable (including the use of supervisory ratings of teacher performance), age, and venue of publication.

Study Size and Design

In one part of her review, Walsh bemoans the lack of experimental research. She then rejects the results of studies with experimental designs because of their smaller sample sizes and

---

little disagreement about the value of having teachers demonstrate their basic skills and subject matter knowledge through either coursework or testing.

cites almost exclusively non-experimental correlational studies, which – though larger – lack direct controls for the variables of interest and must rely on statistical manipulations of data to account, indirectly, for these other influences. This kind of correlational research is, of course, legitimate for staking out broad possibilities in relationships among variables, but it has its own limitations. Many of the more carefully controlled experimental designs can in fact offer more solid evidence about effects, because the "treatment" they are studying is known and the samples can be better controlled than is true for large correlational studies that use proxies and statistical controls rather than direct observation of the phenomena of interest. Medical research, for example, typically uses small sample experimental research as the basis for establishing the possibilities of effects, while using large correlational studies as rough indicators of possible relationships that then require further examination. Single case studies of clinical findings are part of the medical research base along with small carefully controlled experiments, larger clinical trials, and correlational studies looking at broad tendencies.

The usefulness of small, experimental and quasi-experimental studies – including those that Walsh cites and sometimes dismisses (and other times embraces, depending on her reading of and agreement with the findings) – is not in the definitiveness of their individual findings but in their contribution to a larger body of work from which a preponderance of evidence can be examined. Although medical researchers generally consider larger correlational studies to comprise a weaker source of definitive evidence about effects than smaller experimental designs, they recognize that mixed methods of research serve complementary purposes.

Of course, one of the reasons correlational studies must be interpreted with caution is that there is always the question of what direction the correlations may point, sometimes referred to as "reverse causation." There is also the problem that variables in these studies are frequently

crude proxies for the actual measures of interest and may either fail to capture the intended construct or in fact be reflecting the influences of other unmeasured variables. As noted above, many of the variables that can arguably be said to reflect constructs of interest are highly correlated with one another. Furthermore, many of the variables of interest are not well-represented in large data sets. Thus it is critical to represent in any review of research a range of studies that can tease apart the different relationships of interest with a range of measures.

Level of Aggregation

Similarly, studies at different levels of aggregation provide different kinds of insights about the phenomena under study. No study or data set is without limitations, and interpretations should reflect these. In building a corpus of research on any topic, a wide array of research strategies and levels of analyses are used.

It is true that the size of measured effects of different variables can vary at different levels of the system; however, it is not always clear in which way the bias will operate. Often, the general direction of the results holds at different levels of the system, even if effect sizes differ. For example, in their Alabama study, Ferguson and Ladd (1996) found the effects on student achievement of teachers' test scores, masters degrees, and experience held at both the district and school levels in terms of both significance and directionality. There are pros and cons of both kinds of analyses. On the one hand, disaggregated data can exhibit greater measurement error. On the other hand some analysts have argued that omitted variables may bias the coefficients of school input variables upward when data are aggregated to the district or state level (Hanushek, Rivkin, & Taylor, 1995). However, this generalization does not always prove true. For example, although Summers and Wolfe (1975) found that selectivity ratings of each teacher's undergraduate institution were important in explaining 6th grade students' achievement when

examined at the individual teacher level, this relationship disappeared with they aggregated the college ratings and other school inputs into school-level averages. This contradicts the assumption about the usual direction of aggregation bias.

Of course, omitted variables can bias results at any level of the system. Sometimes, especially when the goal of a study is to evaluate broad trends and policy influences, it is important to have data aggregated and analyzed at multiple levels. For interpreting the weight of evidence on a particular issue, the most important question is whether consistent results are found at different levels of aggregation. So, for example, Walsh (p. 6) cites highly aggregated data (Coleman, 1966; Ferguson, 1991; Strauss & Sawyer, 1986) as well as less aggregated data (Hanushek, 1992; Summers & Wolfe, 1977) on the question of the influences of verbal ability. Similarly, the studies examined here reveal influences of measures of teacher education and certification on student achievement at the levels of state (Darling-Hammond, 2000c), school district (Ferguson, 1991; Ferguson & Ladd, 1996; Strauss & Sawyer, 1986), school (Ferguson & Ladd, 1996; Fetler, 1999), and individual teacher (Goldhaber & Brewer, 2000; Hawk, Coble, & Swanson, 1985; Monk, 1994).

Measures for Assessing Teacher Performance

Walsh argues that studies using various ratings of student performance other than student achievement test scores should be discounted, noting that supervisory ratings "can be too subjective to measure teacher quality accurately" (p. 20). As support for this, she cites in her appendix a review of research on teacher evaluation I conducted with colleagues at the RAND Corporation (Darling-Hammond, Wise, & Pease, 1983). While her statement of why I cited the review in another article is completely inaccurate,[23] she is partly right when she notes that

[23] In her separately-published appendix, Walsh states that, "In 1999, Darling-Hammond summarized the main point of this article as a call for using student achievement as the measure of teacher quality." In fact, in Darling-

teacher evaluations by principals and other school-based supervisors have been found to lack strong reliability. Our study of evaluation practices noted that this has been a function of principals' lack of time, inadequate expertise for evaluating all teaching situations, insufficient evaluation training, and inappropriate instrumentation. However, this critique does not extend to ratings of performance that are based on structured observations conducted by trained, expert raters that have been developed and demonstrated to have high reliability. A number of the studies Walsh dismisses use systematic ratings systems by trained observers (e.g. Ferguson & Womack, 1993; Guyton & Farokhi, 1987). The extent to which ratings of performance should be considered or discounted depends on who conducts the rating process, with what training and instrumentation, under what conditions, and with what efforts to enhance reliability.

Age of Studies

The age of studies is also a legitimate but not determinative issue. Studies do not become invalid merely because they are old. While Walsh argues that many older studies using large data sets lacked certain kinds of variables as controls, this does not stop her from citing many of these studies for propositions with which she agrees. More important, the designs of some older studies are at least as strong as some of the more recent studies, and weak studies exist now as then. There is not a strong relationship between study vintage and quality. It is certainly true that teacher education programs and certification requirements have changed over time, so that inferences from studies conducted in one era do not automatically generalize to others; the extent to which one can learn something of use from a study depends on how well the variables are defined and on a knowledge of their relevance to more recent conditions as well as on the strengths and limits of its methodology.

Hammond (1999), I cited this review for an entirely different point. I cited it for the proposition that "Teachers' abilities to structure material, ask higher order questions, use student ideas, and probe student comments have also

Vintage does influence the prevalence of studies of certain kinds. With respect to studies of the effects of teacher education and certification, a large number of studies were conducted in the high-demand era of the 1960s and '70s when there was great variability in entry pathways and much interest in the topic. It is also true that federal funding for educational research was substantially larger before 1980 than it was during the severe budget cuts of that decade. In addition, in times of relatively low demand, like most of the 1980s, virtually all teachers were certified and there was too little variability to find effects of this variable in large-scale studies. Few studies were concerned with these issues and few data sets had measures of teacher education variables. Interest and data on this topic have just begun to return in the 1990s. Those who are interested in the extent to which – and the ways in which – different kinds of preparation may matter for teacher performance and student learning can and should be informed by earlier studies where they are applicable to the questions under study.

Publication Venue

Although Walsh is incorrect in her statement that dissertations are not retrievable (there are library systems for doing so, if sometimes less than convenient), it is legitimate to suggest that the kind of review they have received is often more variable, and may be less strenuous depending on the university and department, than for many peer-reviewed journals. There are certainly some universities whose dissertation review process is more rigorous than some journals, but the reverse is also certainly true. The same variability in review stringency is true for conference papers and technical reports. However, Walsh herself cites a substantial number of unreviewed papers in support of various positions she takes. There are different schools of thought about how to treat these papers in reviews. Some would argue, as does Veenman (1984), a reviewer cited by Walsh, that the use of all identified studies is justified for a review

---

been found to be important variables in what students learn."

that seeks to delineate global trends where large numbers of findings are similar (p. 166). Others would argue that papers that have not been published with peer review should be used only when the review includes a critique of each study's methods. Others might argue, as Walsh does (at least rhetorically if not in practice), that such studies should be excluded from consideration. I accept the point that it is a useful common ground to rely on research published in peer-reviewed journals, and I restrict the analysis in this paper to those studies. Even with this criterion, there is substantial evidence to be weighed and discussed.

Who is Affected by this Debate?

The critical issue here is not the protection of researchers' reputations or the turf of schools of education but the protection of students, especially low-income students and students of color who are disproportionately taught by unprepared and uncertified teachers. As Walsh's paper shows in her references to data on the disparities in access to qualified teachers for students in Baltimore, the children most affected by these arguments are economically and educationally disadvantaged children in central cities who are substantially abandoned by the funding and hiring protections that should operate to provide a foundation for their education. These are the students whose education is most undermined by their lack of access to teachers who have the knowledge and skills to ensure that they learn to the new high standards the society and the state demand.

What the statistics on the lack of certified teachers actually mean on the ground is that many of Baltimore's most educationally vulnerable children – most of them African American – are taught in their elementary school years by teachers who have had no training in how to teach them to read, much less to develop other basic and higher order skills they must have to succeed in school and life. When they fail to learn, they begin the tortuous process of educational failure

that will end for many of them in dropping out or being unable to pass the state tests that would grant them a diploma. This then launches a life spent either in a marginal part of the economy that barely yields subsistence wages or, as is true for more than 50% of high school dropouts, in the inability to gain any job at all. In today's economy, these young people are fated to become part of the growing criminal justice system, as incarceration is increasingly linked to inadequate education. More than half of the growing number of inmates in the United States are functionally illiterate and cannot gain access to today's labor market. This is not unrelated to the fact that so many low-income students have been taught by teachers who never learned how to teach them to read.

Illogical Policy Conclusions

The disparities in access to qualified teachers in Maryland are a function of a state school finance system that has underfunded Baltimore's schools for decades, along with inadequate incentives – for example, service scholarships, forgivable loans, and recruitment attractions like salaries and housing assistance – to encourage individuals to acquire strong training and then teach in high-need fields and locations. The Abell Foundation report does not argue for more equitable funding for the schools that serve Maryland's poor and minority students or for stronger incentives to attract well-prepared teachers to these schools. In fact, the report cites approvingly a paper prepared to stave off an equity lawsuit in Maryland (Hanushek, 1996b) which argues against district investments in smaller class sizes or higher salaries in Baltimore, asserting that "Baltimore City would not benefit from additional resources as much as it could benefit by better school management."[24] The Abell Foundation report argues that the enormous disparities in resources and qualified teachers between Baltimore and other districts are not a

[24] Cited in the separately-published appendix entry 88, p.50.

problem because teacher certification does not mean anything, and that in fact the solution is to do away with certification altogether.

The outcome of Walsh's argument, were it to be successful in the policy community, would be continued inequality in funding, depressed salaries for teaching in high-need areas, continued lack of access for poor children to a stable teaching force of well-qualified teachers by any definition, and tragic loss of life for students who are underserved.

To be sure, certification is but a proxy for the subject matter knowledge and knowledge of teaching and learning embodied in various kinds of coursework and in the evidence of ability to practice contained in supervised student teaching. It is true that certification is a relatively crude measure of teachers' knowledge and skills, since the standards for subject matter and teaching knowledge embedded in certification have varied across states and over time, are differently measured, and are differently enforced from place to place. The quality of preparation in both university programs and other alternatives has varied as well, although a number of states have made substantial recent headway in strengthening teachers' preparation and reducing this variability. Given the crudeness of the measure, it is perhaps remarkable that many studies have found significant effects of teacher certification.

This does not mean that we should be sanguine about certification policies. There are questions about the quality of tests, courses, and institutions that are the subject of study and action across the country (see, for example, Darling-Hammond, Wise, & Klein, 1995). The answer to flaws that may be perceived, however, is not to eliminate or undermine the pathways that enable and require teachers to gain knowledge and students to have access to teachers who have the knowledge they need. If teacher knowledge and skill about both content and how to teach it is important, as substantial evidence suggests it is, the most sensible policy goal is to

work to improve preparation opportunities and certification standards so that they increasingly approximate what teachers need to know and do in order to be successful with diverse students.

As Levin (1980) notes, certification is a critically important exercise in the economics of information that should be a target of continual improvement:

> (T)he facts that we expect the schools to provide benefits to society that go beyond the sum of those conferred upon individual students, that it is difficult for many students and their parents to judge certain aspects of teacher proficiency, and that teachers cannot be instantaneously dismissed, mean that somehow the state must be concerned about the quality of teaching. It cannot be left only to the individual judgments of students and their parents or the educational administrators who are vested with managing the schools in behalf of society. The purpose of certification of teachers and accreditation of the programs in which they received their training is to provide information on whether teachers possess the minimum proficiencies that are required from the teaching function. Because this is an exercise in the provision of information, it is important to review the criteria for setting out how one selects the information that is necessary to make a certification or accreditation decision (p. 7).

Conclusion

Kate Walsh has dismissed or misreported much of the existing evidence base in order to argue that teacher education makes no difference to teacher performance or student learning and that students would be better off without state efforts to regulate entry into teaching or to provide supports for teachers' learning. While she argues for recruiting bright people into teaching (and who could disagree with that?), her proposals offer no incentives for attracting individuals into teaching other than the removal of preparation requirements. While this proposal is couched as the elimination of "barriers" to teaching, evidence suggests that lack of preparation actually contributes to high attrition rates and thereby becomes a disincentive to long-term teaching commitments and to the creation of a stable, high ability teaching force. Lack of preparation also contributes to lower levels of learning, especially for those students who most need skillful teaching in order to succeed.

The evidence from research presented here and elsewhere makes clear that the policies Walsh endorses could bring harm to many children, especially those who are already least well

served by the current system. Those who make such arguments for eliminating one of the few protections these children have should bear the burden of proof for showing how what they propose could lead to greater equity and excellence in American schools.

References

Andrew, M. & Schwab, R.L. (1995). Has reform in teacher education influenced teacher performance? An outcome assessment of graduates of eleven teacher education programs. *Action in Teacher Education, 17*: 43-53.

Andrews, J.W., Blackmon, C.R., & Mackey, J.A. (1980). Preservice performance and the National Teacher Examinations. *Phi Delta Kappan, 61(5)*: 358-359.

Ashton, P. & Crocker, L. (1987). Systematic study of planned variations: The essential focus of teacher education reform. *Journal of Teacher Education,* 2-8.

Ayers, J.B., & Qualls, G.S. (1979). Concurrent and predictive validity of the National Teacher Examinations. *Journal of Educational Research, 73 (2)*: 86-92.

Begle, E.G. (1979). *Critical Variables in Mathematics Education: Findings from a Survey of the Empirical Literature.* Washington, DC: Mathematical Association of American and National Council of Teachers of Mathematics.

Begle, E.G. (1972). *Teacher Knowledge and Student Achievement in Algebra, Report 9.* Stanford University, California School Mathematics Study Group. Washington, DC: National Science Foundation.

Begle, E.G. & Geeslin, W. (1972). *Teacher effectiveness in mathematics instruction. National Longitudinal Study of Mathematical Abilities Reports No. 28.* Washington, DC: Mathematical Association of America and National Council of Teachers of Mathematics.

Berliner, D.C. (1986). In pursuit of the expert pedagogue, *Educational Researcher* (August/September): 5-13.

Bliss, T. (1992). Alternative certification in Connecticut: Reshaping the profession. *Peabody Journal of Education, 67(3)*: 35-54.

Bolich, A.M. (2001). *Reduce Your Losses: Help New Teachers Become Veteran Teachers.* Atlanta, GA: Southern Regional Education Board.

Borko, H. & Mayfield, V. (1995). The roles of the cooperating teacher and university supervisor in learning to teach. *Teaching and Teacher Education, 11(5)*: 501-518.

Bowles, S., & Levin, H.M. (1968). The determinants of scholastic achievement- An appraisal of some recent evidence. *Journal of Human Resources, 3*: 3-24.

Bradshaw, L. & Hawk, P. (1996). *Teacher Certification: Does It Really Make a Difference in Student Achievement?* Greenville, NC: Eastern North Carolina Consortium for Assistance and Research in Education.

Bryk, A.S. & Lee V.E. (1992). Are politics the problems and markets the answer? An essay review of "Politics, markets and America's schools." *Economics of Education Review, 11(4)*: 439-451.

Byrne, C.J. (1983). *Teacher Knowledge and Teacher Effectiveness: A Literature Review, Theoretical Analysis and Discussion of Research Strategy*. Paper presented at the meeting of the Northwestern Educational Research Association, Ellenville, NY.

Carroll, J.B. (1975). *The Teaching of French as a Foreign Language in Eight Countries*. New York: John Wiley and Sons.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). *Equality of Educational Opportunity*. Washington, DC: U.S. Government Printing Office.

Coleman S. J., Hoffer T., and Kilgore, S. (1982). Cognitive outcomes in public and private schools. *Sociology of Education, 55* (2-3): 65-76.

Crow, N., Knowles, J., & Bullough, R. (1989). Teacher self-concept and student culture in the first year. *Teachers College Record, 91(2)*: 209-233.

Darling-Hammond, L. (1992). Teaching and knowledge: Policy issues posed by alternate certification for teachers. *Peabody Journal of Education, 67(3)*: 123-154.

Darling-Hammond, L. (1997). *Doing What Matters Most: Investing in Quality Teaching*. NY: National Commission on Teaching and America's Future, Teachers College, Columbia University.

Darling-Hammond, L. (1999). *Teacher Quality and Student Achievement: A Review of State Policy Evidence*. Seattle: Center for the Study of Teaching and Policy, University of Washington.

Darling-Hammond, L. (2000a). Reforming teacher preparation and licensing: Debating the evidence. *Teachers College Record, 102*, (1): 28-56.

Darling-Hammond, L. (2000b). *Solving the Dilemmas of Teacher Supply, Demand, and Standards: How We Can Ensure a Competent, Caring, and Qualified Teacher for Every Child*. NY: National Commission on Teaching and America's Future.

Darling-Hammond, L. (2000c). Teacher quality and student achievement. *Education Policy Analysis Archives, 8(1)*: http://epaa.asu.edu/epaa/v8n1.html

Darling-Hammond, L., Berry, B., & Thoreson, A. (2001). Does Teacher Certification Matter? Evaluating the Evidence. *Educational Evaluation and Policy Analysis, 23(1)*: 57-77.

67 BEST COPY AVAILABLE

Darling-Hammond, L., Hudson, L., & Kirby, S. (1989). *Redesigning Teacher Education: Opening the Door for New Recruits to Science and Mathematics Teaching*. Santa Monica: The RAND Corporation.

Darling-Hammond, L., Wise, A.E. & Pease, S.R. (1983). Teacher evaluation in the organizational context: a review of the literature. *Review of Educational Research, 53*: 285-237.

Denton, J.J., & Peters, W.H. (1988). *Program Assessment Report: Curriculum Evaluation of a Non-traditional Program for Certifying Teachers*. Texas A&M University, College Station, TX.

Druva, C.A., & Anderson, R.D. (1983). Science teacher characteristics by teacher behavior and by student outcome: A meta-analysis of research. *Journal of Research in Science Teaching, 20(5)*: 467-479.

Evertson, C., Hawley, W., & Zlotnick, M. (1985). Making a difference in educational quality through teacher education. *Journal of Teacher Education, 36* (3), 2-12.

Feistritzer, C.E. (1984). *The Making of a Teacher*. Washington, DC: National Center for Education Information.

Ferguson, R.F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation, 28(2)*: 465-498.

Ferguson, R.F. & Ladd, H.F. (1996). How and why money matters: An analysis of Alabama schools. In Helen Ladd (ed.) *Holding Schools Accountable*, pp. 265-298. Washington, DC: Brookings Institution.

Ferguson, P. & Womack, S.T. (1993). The impact of subject matter and education coursework on teaching performance. *Journal of Teacher Education, 44* (1): 55-63.

Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives, 7(9)*: http://epaa.asu.edu/epaa/v7n9.html.

Gitomer, D.H., Latham, A.S., & Ziomek, R. (1999). *The Academic Quality of Prospective Teachers: The Impact of Admissions and Licensure Testing*. Princeton, NJ: Educational Testing Service.

Goldhaber, D.D. & Brewer, D. J. (1998, October). When should we reward degrees for teachers? Phi Delta Kappan: 134-138.

Goldhaber, D.D. & Brewer, D.J. (2000). Does teacher certification matter? High school certification status and student achievement. *Educational Evaluation and Policy Analysis, 22*: 129-145.

Grey, L., Cahalan, M., Hein, S., Litman, C., Severynse, J., Warren, S., Wisan, G., & Stowe, P. (1993). *New Teachers in the Job Market: 1991 Update.* Washington, DC: U. S. Department of Education, Office of Educational Research and Improvement.

Greenwald, R., Hedges, L.V., & Laine, R.D. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*: 361-396.

Guyton, E. & Farokhi, E. (1987). Relationships among academic performance, basic skills, subject matter knowledge and teaching skills of teacher education graduates. *Journal of Teacher Education* (Sept-Oct.): 37-42.

Haney, W., Madaus, G., & Kreitzer, A. (1987). Charms talismanic: testing teachers for the improvement of American education. In E.Z. Rothkopf (Ed.) *Review of Research in Education, 14*: 169-238. Washington, DC: American Educational Research Association.

Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives, 8* (41): http://epaa.asu.edu/epaa/v8n41/

Hanushek, E. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *The American Economic Review 61(2)*: 280-288.

Hanushek, E. (1992). The trade-off between child quantity and quality. *Journal of Political Economy, 100*: 84-117.

Hanushek, E.A., Rivkin, S.G., &Taylor, L.L. (1995). *Aggregation Bias and the Estimated Effects of School Resources.* Rochester, NY: University of Rochester, Center for Economic Research.

Hanushek, E. (1996b). *School Resources and achievement in Maryland.* Baltimore, MD: Maryland State Department of Education.

Hawk, P., Coble, C.R., & Swanson, M. (1985). Certification: It does matter. *Journal of Teacher Education, 36(3)*: 13-15.

Hellfritzch, A.G. (1945). A factor analysis of teacher abilities. *Journal of Experimental Education, 14*: 166-169.

Ingersoll, R. (1998). The problem of out-of-field teaching. *Phi Delta Kappan,* (June 1998): 773-776.

Jelmberg, J. (1995). College-based teacher education versus state-sponsored alternative programs. *Journal of Teacher Education, 47*(1), 60-66. (Jan-Feb 1996).

LaDuke, D.V. (1945). The measurement of teaching ability. *Journal of Experimental Education, 14*: 75-100.

Lanier, J. and J. Little. (1986). Research on Teacher Education. In M. Wittrock (ed.), *Handbook of Research on Teaching, Third Edition*. New York: Macmillan.

Lee, V.E. & Byrk, A.S. (1988) Curriculum tracking as mediating the social distribution of high school achievement. *Sociology of Education, 61*: 78-94.

Lee, V.E., Dedrick, R.F., & Smith, J.B. (1991) The effect of the social organization of schools on teachers' self efficacy and satisfaction. *Sociology of Education, 64*: 190-208.

Levin, H. M. (1980). Teacher certification and the economics of information. *Educational Evaluation and Policy Analysis, 2 (4)*: 5-18.

Little, J.W. (1999). Organizing schools for teacher learning. In L. Darling-Hammond and G. Sykes (eds.), *Teaching as the Learning Profession*, pp. 233-262. San Francisco: Jossey-Bass.

Lutz, F.W. & Hutton, J.B. (1989). Alternative teacher certification: Its policy implications for classroom and personnel practice. *Educational Evaluation and Policy Analysis, 11(3)*: 237-254.

Miller, J.W., McKenna, M.C., & McKenna, B.A. (1998). A comparison of alternatively and traditionally prepared teachers. *Journal of Teacher Education, 49(3)*: 165- 176.

Mitchell, N. (1987). *Interim Evaluation Report of the Alternative Certification Program (REA87-027-2)*. Dallas, TX: DISD Department of Planning, Evaluation, and Testing.

Monk, D. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 12(2)*: 125-142.

Monk, D. & King, J. (1994). Multi-level teacher resource effects on pupil performance in secondary mathematics and science. In R.G. Ehrenberg (ed.), *Choices and Consequences*. ILR Press, Ithaca, NY.

Murnane, R.J. (1985). *Do Effective Teachers have Common Characteristics: Interpreting the Quantitative Research Evidence*. Paper presented at the National Research Council Conference on Teacher Quality in Science and Mathematics, Washington, DC

Murnane, R.J. (1983). Understanding the sources of teaching competence: Choices, skills and the limits of training. *Teachers College Record, 84(3)*: 564-569.

National Center for Education Statistics (NCES) (1985). *The Condition of Education, 1985*. Washington, DC. U.S. Department of Education.

National Center for Education Statistics (NCES) (1997). *America's Teachers: Profile of a Profession*. Washington, DC: U.S. Department of Education.

National Center for Education Statistics (NCES) (2000). *Digest of Education Statistics, 1999*. Washington, DC: U.S. Department of Education.

National Commission on Teaching and America's Future (1996). *What Matters Most: Teaching for America's Future.* New York: Author.

National Reading Panel (2000). *Teaching Children to Read: An Evidence-based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction.* Washington, DC: National Institute of Child Health and Human Development.

Natriello, G. Zumwalt, K., Hansen, A. & Frisch, A. (1990). *Characteristics of Entering Teachers in New Jersey.* Paper presented at the 1988 Annual Meeting of the American Educational Research Association.

Olsen, D.G. (1985). The quality of prospective teachers: Education vs. noneducation graduates. *Journal of Teacher Education, 36(5)*: 56-59.

Quirk, T.J., Witten, B.J., & Weinberg, S.F. (1973). Review of studies of concurrent and predictive validity of the National Teacher Examinations. *Review of Educational Research, 43*: 89-114.

Raymond, M., Fletcher, S., & Luque, J., 2001, *Teach for America: An Evaluation ofTteacher Differences and Student Outcomes in Houston, Texas.* CREDO, The Hoover Institution, Stanford University: www.rochester.edu/credo

Rostker, L.E. (1945). The measurement of teaching ability. *Journal of Experimental Education, 14*: 5-51.

Schalock, D. (1979). Research on teacher selection. In D.C. Berliner (ed.), *Review of Research in Education (vol. 7)*, Washington, DC: American Educational Research Association.

Shotel, J.R. (Summer 1998). *Does NCATE Make a Difference? Quality in Teacher Education.* Washington, DC: George Washington University.

Skinner, W.A. (1947). *An Investigation of Factors Useful in Predicting Teaching Ability.* University of Manchester. Master of Education thesis.

Stanford Research International (SRI) (2000). *The Status of the Teaching Profession, 2000: An Update to the Teaching and California's Future Task Force.* Santa Cruz, CA: The Center for the Future of Teaching and Learning.

Soar, R.S., Medley, D.M., and Coker, H. (1983). Teacher evaluation: A critique of currently used methods. *Phi Delta Kappan, 65(4)*: 239-246.

Stoddart, Trish (1992). An alternate route to teacher certification: Preliminary findings from the Los Angeles Unified School District Intern Program. *Peabody Journal of Education, 67(3).*

Strauss, R.P. & Sawyer, E.A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review, 5(1)*: 41-48.

Summers, A.A., & Wolfe, B.L. (1975; 1977). Which School Resources Help Learning? Efficiency and Equality in Philadelphia Public Schools. *The American Economic Review, 67(4)*: 639-652.

Texas Center for Educational Research (2000). *The Cost of Teacher Turnover*. Austin, TX: Texas State Board for Teacher Certification (SBEC).

Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research, 54*: 143-178.

Vernon, P.E. (1965). Personality factors in teacher trainee selection. *British Journal of Education Psychology (35)*: 140-149.

Wenglinsky, H. (2000). *How teaching matters: Bringing the classroom back into discussions of teacher quality*. Princeton, NJ: Educational Testing Service.

Wilson, S., Floden, R., & Ferrini-Mundy (2001). *Teacher Preparation Research: Current Knowledge, Gaps, and Recommendations*. University of Washington: Center for the Study of Teaching and Policy.

Wright, David P., Michael McKibbon, & Priscilla Walton (1987). *The Effectiveness of the Teacher Trainee Program: An Alternate Route into Teaching in California*. California Commission on Teacher Credentialing.

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

REPRODUCTION BASIS

☑ This document is covered by a signed "Reproduction Release (Blanket) form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").